PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-105347
(43)Date of publication of application : 24.04.1998

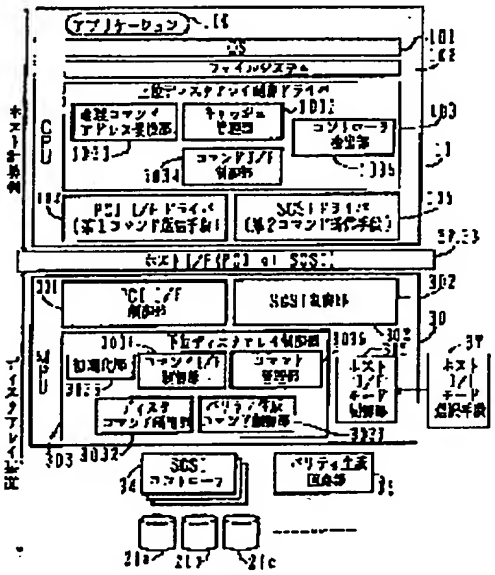
(51)Int.Cl. G06F 3/06
G06F 3/06

(21)Application number : 08-262147
(22)Date of filing : 02.10.1996

(71)Applicant : HITACHI LTD
(72)Inventor : MATSUNAMI NAOTO
OEDA TAKASHI
KANEDA TAISUKE
ARAKAWA TAKASHI
YAGISAWA IKUYA
TAKANO MASAHIRO

(54) DISK ARRAY CONTROL SYSTEM

(57)Abstract:
PROBLEM TO BE SOLVED: To provide a high-performance disk array control system whose performance is not limited even when an inexpensive micro processing unit(MPU) is used for a disk array controller.
SOLUTION: The CPU 10 of a host computer is equipped with a host disk array control driver 103 which generates a disk command from a logical command to a disk array device and also generates a parity generation command. The disk array controller is equipped with an MPU 30 which has lower throughput than a CPU that executes a control program in the disk array controller, a SCSI controller 34, and a parity generating circuit 35. The MPU 30 is equipped with a slave disk array control part 303 which receives the disk command and parity generation command issued by the host disk array control driver 103 and sends them to disks and the parity generating circuit 35.



LEGAL STATUS

[Date of request for examination] 06.03.2000
[Date of sending the examiner's decision of rejection] 24.09.2003
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
[Date of final disposal for application]
[Patent number]
[Date of registration]
[Number of appeal against examiner's decision of rejection]
[Date of requesting appeal against examiner's decision of rejection]
[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（Ａ）**　(11)特許出願公開番号

特開平10－105347

(43)公開日　平成10年(1998)４月24日

(51)Int.Cl.[6]	識別記号	ＦＩ	
Ｇ０６Ｆ　3/06	５４０	Ｇ０６Ｆ　3/06	５４０
	３０５		３０５Ｃ

審査請求　未請求　請求項の数９　ＯＬ　（全 38 頁）

(21)出願番号　特願平8－262147

(22)出願日　平成８年(1996)10月２日

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地

(72)発明者　松並　直人
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内

(72)発明者　大枝　高
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内

(72)発明者　兼田　泰典
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内

(74)代理人　弁理士　春日　譲

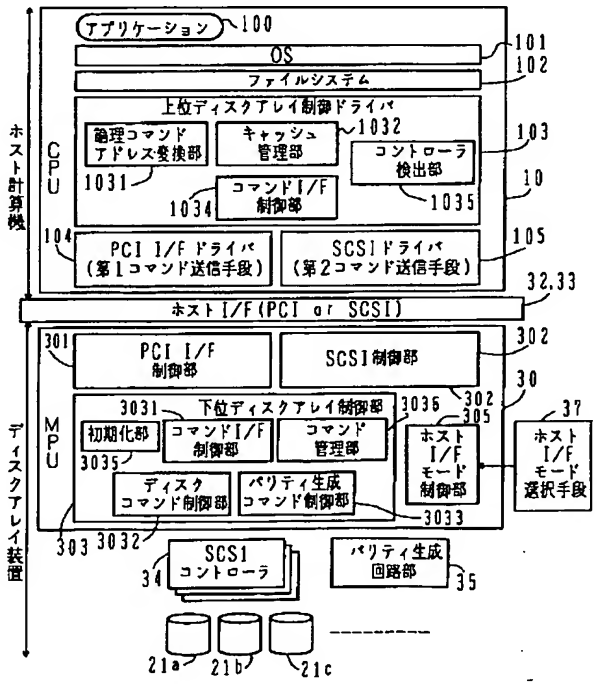
最終頁に続く

(54)【発明の名称】　ディスクアレイ制御システム

(57)【要約】　（修正有）

【課題】ディスクアレイコントローラに低価格なＭＰＵを使用しても性能が制限されない高性能なディスクアレイ制御システムを提供する。

【解決手段】ホスト計算機のＣＰＵ１０は、ディスクアレイ装置への論理コマンドからディスクコマンドを生成し、かつ、パリテイ生成コマンドを生成する上位ディスクアレイ制御ドライバ１０３を備えている。ディスクアレイコントローラは、ディスクアレイコントローラの内部の制御プログラムを実行するＣＰＵよりも処理能力の小さいＭＰＵ３０と、ＳＣＳＩコントローラ３４と、パリテイ生成回路３５とを備える。ＭＰＵ３０は、上位ディスクアレイ制御ドライバ１０３が発行したディスクコマンド及びパリテイ生成コマンドを受信してそれらをディスクやパリテイ生成回路３５に対して発行する下位ディスクアレイ制御部３０３を備えている。



【特許請求の範囲】
【請求項１】 プログラムを実行するＣＰＵを有するホスト計算機と、
このホスト計算機にホストインターフェースを介して接続されるとともに、複数のディスクからなるディスクアレイと、このディスクアレイを制御するディスクアレイコントローラとを有するディスクアレイ装置から構成され、
上記ＣＰＵは、上記ディスクアレイ装置への論理コマンドから上記ディスクアレイを構成するディスクへのディスクコマンドを生成し、かつ、冗長データを生成するための冗長データ生成コマンドを生成する上位ディスクアレイ制御手段を備え、
上記ディスクアレイコントローラは、
上記ディスクアレイコントローラの内部の制御プログラムを実行する上記ＣＰＵよりも処理能力の小さいＭＰＵと、
上記ディスクアレイを構成するディスクを接続するディスクインターフェースと、
上記ディスクアレイに格納する冗長データを生成する冗長データ生成手段とを備え、
上記ＭＰＵは、上記上位ディスクアレイ制御手段が発行した上記ディスクコマンド及び冗長データ生成コマンドを受信し、上記ディスクアレイを構成するディスクにディスクコマンドを発行し、また、上記冗長データ生成手段に冗長データ生成コマンドを発行する下位ディスクアレイ制御手段を備えることを特徴とするハイブリッドアレイ構成のディスクアレイ制御システム。
【請求項２】 プログラムを実行するＣＰＵを有するホスト計算機と、
このホスト計算機にホストインターフェースを介して接続されるとともに、複数のディスクからなるディスクアレイと、このディスクアレイを制御するディスクアレイコントローラとを有するディスクアレイ装置から構成され、
上記ディスクアレイ装置の上記ホストインターフェースは、上記ホスト計算機の内部システムバスに直結するカードエッジを有する構造の第１ホストインターフェースと、
上記ホスト計算機とケーブルを介して接続する構造の第２ホストインターフェースとから構成され、
上記ディスクアレイ装置は、上記第１ホストインターフェースと上記第２ホストインターフェースのいずれか一方を選択するホストインターフェースモード選択手段を備え、内蔵型と外付け型を選択できることを特徴とするディスクアレイ制御システム。
【請求項３】 請求項２記載のディスクアレイ制御システムにおいて、
上記ＣＰＵは、さらに、
上記第１ホストインターフェースに上記ＣＰＵが生成したコマンドを送信し、上記ディスクアレイコントローラからの終了報告を受信する第１コマンド送信制御手段と、
上記第２ホストインターフェースに上記ＣＰＵが生成したコマンドを送信し、ディスクアレイコントローラからの終了報告を受信する第２コマンド送信制御手段と、
上記第１若しくは第２コマンド送信制御手段を選択するコマンド送信選択手段を備え、
上記ディスクアレイコントローラは、さらに、
上記第１ホストインターフェースから上記ＣＰＵが生成したコマンドを受信し、その終了を上記ホスト計算機に報告する第１コマンド受信制御手段と、
上記第２ホストインターフェースから上記ＣＰＵが生成したコマンドを受信し、その終了をホストに報告する第２コマンド受信制御手段と、
上記ホストインターフェースモード選択手段により決定した上記第１若しくは第２ホストインターフェースに応じた上記第１若しくは第２コマンド受信制御手段を選択するコマンド受信選択手段とを備えたことを特徴とするディスクアレイ制御システム。
【請求項４】 請求項２記載のディスクアレイ制御システムにおいて、
上記ホストインターフェースモード選択手段が、上記第２ホストインターフェースを選択した際には、
上記第２ホストインターフェースと上記ホスト計算機はケーブルで接続し、
上記ホスト計算機の外部に設けたディスクアレイ筐体に上記ディスクアレイコントローラ及び上記ディスクアレイを搭載し、
上記ディスクアレイ筐体は、上記第１ホストインターフェースのカードエッジを接続する構造のコネクタと、電源と、クロック信号生成手段を備え、
上記ディスクアレイコントローラは、第１ホストインターフェースのカードエッジで上記コネクタに接続し、上記コネクタを介して上記電源から電力を供給し、上記クロック信号生成手段からクロック信号を供給することを特徴とするディスクアレイ制御システム。
【請求項５】 請求項４記載のディスクアレイ制御システムにおいて、
上記ディスクアレイ筐体は、上記ホストインターフェースモード選択手段と接続するコネクタを有し、
上記ホストインターフェースモード選択手段は、上記ディスクアレイ筐体のコネクタと接続した際には第２ホストインターフェース制御手段を選択することを特徴とするディスクアレイ制御システム。
【請求項６】 プログラムを実行するＣＰＵを有するホスト計算機と、
このホスト計算機にホストインターフェースを介して接続されるとともに、複数のディスクからなるディスクアレイと、このディスクアレイを制御するディスクアレイ

コントローラとを有するディスクアレイ装置から構成され、
上記ＣＰＵ若しくは上記ＭＰＵは、上記ディスクアレイ装置への論理コマンドから上記ディスクアレイを構成するディスクへのディスクコマンドを生成し、かつ、冗長データを生成するための冗長データ生成コマンドを生成する上位ディスクアレイ制御手段を備え、
上記ディスクアレイコントローラは、
上記ディスクアレイコントローラの内部の制御プログラムを実行する上記ＣＰＵよりも処理能力の小さいＭＰＵと、
上記ディスクアレイを構成するディスクを接続するディスクインターフェースと、
上記ディスクアレイに格納する冗長データを生成する冗長データ生成手段とを備え、
上記ＭＰＵは、上記上位ディスクアレイ制御手段が発行した上記ディスクコマンド及び冗長データ生成コマンドを受信し、上記ディスクアレイを構成するディスクにディスクコマンドを発行し、また、上記冗長データ生成手段に冗長データ生成コマンドを発行する下位ディスクアレイ制御手段を備え、
さらに、上記ディスクアレイコントローラは、上記ＣＰＵ若しくは上記ＭＰＵが備える上記上位ディスクアレイ制御手段を選択するディスクアレイモード選択手段を備え、ハイブリッドアレイ構成とハードアレイ構成を選択できることを特徴とするディスクアレイ制御システム。
【請求項７】 プログラムを実行するＣＰＵを有する複数のホスト計算機と、
これらのホスト計算機にホストインターフェースを介して接続されるとともに、複数のディスクからなるディスクアレイと、このディスクアレイを制御するディスクアレイコントローラとを有するディスクアレイ装置から構成され、
上記複数のホスト計算機のそれぞれの上記ＣＰＵは、上記ディスクアレイ装置への論理コマンドから上記ディスクアレイを構成するディスクへのディスクコマンドを生成し、かつ、冗長データを生成するための冗長データ生成コマンドを生成する上位ディスクアレイ制御手段を備え、
上記ディスクアレイコントローラは、
上記ディスクアレイコントローラの内部の制御プログラムを実行する上記ＣＰＵよりも処理能力の小さいＭＰＵと、
上記ディスクアレイを構成するディスクを接続するディスクインターフェースと、
上記ディスクアレイに格納する冗長データを生成する冗長データ生成手段とを備え、
上記ＭＰＵは、上記上位ディスクアレイ制御手段が発行した上記ディスクコマンド及び冗長データ生成コマンドを受信し、上記ディスクアレイを構成するディスクにディスクコマンドを発行し、また、上記冗長データ生成手段に冗長データ生成コマンドを発行する下位ディスクアレイ制御手段を備え、
上記ディスクアレイ装置は、第１の上記ホスト計算機に内蔵され、
上記ディスクアレイ装置の上記ホストインターフェースは、上記第１のホスト計算機の内部システムバスに直結するカードエッジを有する構造の第１ホストインターフェースと、
上記第１のホスト計算機以外の上記ホスト計算機とケーブルを介して接続する構造の第２ホストインターフェースとから構成されることを特徴とするディスクアレイ制御システム。
【請求項８】 プログラムを実行するＣＰＵを有する第１のホスト計算機と、
プログラムを実行するＣＰＵを有する第２のホスト計算機と、
上記第１及び第２のホスト計算機にホストインターフェースを介して接続されるとともに、複数のディスクからなるディスクアレイと、このディスクアレイを制御するディスクアレイコントローラとを有するディスクアレイ装置から構成され、
上記第１の計算機の上記ＣＰＵは、上記ディスクアレイ装置への論理コマンドから上記ディスクアレイを構成するディスクへのディスクコマンドを生成し、かつ、冗長データを生成するための冗長データ生成コマンドを生成する上位ディスクアレイ制御手段を備え、
上記第２の計算機の上記ＣＰＵは、上記ディスクアレイ装置への論理コマンドを上記ディスクアレイコントローラに発行するディスク制御手段を備え、
上記ディスクアレイコントローラは、
上記ディスクアレイコントローラの内部の制御プログラムを実行する上記ＣＰＵよりも処理能力の小さいＭＰＵと、
上記ディスクアレイを構成するディスクを接続するディスクインターフェースと、
上記ディスクアレイに格納する冗長データを生成する冗長データ生成手段とを備え、
上記ＭＰＵは、上記上位ディスクアレイ制御手段が発行した上記ディスクコマンド及び冗長データ生成コマンドを受信し、上記ディスクアレイを構成するディスクにディスクコマンドを発行し、また、上記冗長データ生成手段に冗長データ生成コマンドを発行する下位ディスクアレイ制御手段を備え、
上記ディスクアレイ装置は、第１の上記ホスト計算機に内蔵され、
上記ディスクアレイ装置の上記ホストインターフェースは、上記第１のホスト計算機の内部システムバスに直結するカードエッジを有する構造の第１ホストインターフェースと、

上記第２のホスト計算機とケーブルを介して接続する構造の第２ホストインターフェースとから構成されることを特徴とするディスクアレイ制御システム。

【請求項９】　ホスト計算機にホストインターフェースを介して接続されるとともに、複数のディスクからなるディスクアレイを制御するディスクアレイコントローラにおいて、

上記ディスクアレイコントローラの内部の制御プログラムを実行するＭＰＵと、

上記ディスクアレイを構成するディスクを接続するディスクインターフェースと、

上記ディスクアレイに格納する冗長データを生成する冗長データ生成手段とを備え、

上記ホストインターフェースは、上記ホスト計算機の内部システムバスに直結するカードエッジを有する構造の第１ホストインターフェースと、

上記ホスト計算機とケーブルを介して接続する構造の第２ホストインターフェースとから構成されることを特徴とするディスクアレイコントローラ。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、ディスクアレイを制御するディスクアレイ制御システムに関する。

【０００２】

【従来の技術】ディスクアレイは、複数のディスクを並列に動作させることで、単体のディスクに比べ高速化を実現する技術である。しかし、ディスクをｎ台並べた場合、その故障確率は、ｎ倍に悪化する。

【０００３】そこで、高速化と高信頼化を両立するための技術として、「ＲＡＩＤ（Ｒｅｄｕｎｄａｎｔ Ａｒｒａｙｓ ｏｆ Ｉｎｅｘｐｅｎｓｉｖｅ Ｄｉｓｋｓ）」が知られている。ＲＡＩＤは、例えば、”Ａ Ｃａｓｅ ｆｏｒ Ｒｅｄｕｎｄａｎｔ Ａｒｒａｙｓ ｏｆ Ｉｎｅｘｐｅｎｓｉｖｅ Ｄｉｓｋｓ （ＲＡＩＤ）”； Ｉｎ Ｐｒｏｃ. ＡＣＭ ＳＩＧＭＯＤ, Ｊｕｎｅ １９８８（カリフォルニア大学バークレー校発行）」に記載されている。

【０００４】ＲＡＩＤは、複数のディスクを並列に動作させることで高速制御を実現し、また、パリティと呼ぶ冗長データをパリティディスクと呼ぶ特定のディスクに格納することにより、万一、データを格納する１台のディスクが故障しても、他のディスクとパリティディスクのパリティとから故障したディスクのデータを再現することができ、耐ディスク障害信頼性を高めることができるディスク制御の方法である。

【０００５】ＲＡＩＤは、そのパリティの格納の方法によりレベル１から５がある。レベル４のＲＡＩＤ型ディスクアレイでは、例えば、ディスクが５台あり、４台がデータディスク０～３、１台がパリティディスクとすると、データをディスク０，ディスク１，ディスク２，ディスク３の順番で、ある一定のデータブロック毎に順に分散して格納する。このデータブロック単位のことをストライプと称し、この分散する制御のことをストライピングと称する。

【０００６】ディスク０～３に格納した同一ストライプのデータＤ０～３の排他的論理和（以下ＸＯＲと称する）を計算することで、

Ｐ＝Ｄ０＋Ｄ１＋Ｄ２＋Ｄ３　（ただし、＋はＸＯＲ演算を示す）

パリティＰを生成する事ができる。また、Ｄ０を格納するディスク０が故障した際には、

Ｄ０＝Ｄ１＋Ｄ２＋Ｄ３＋Ｐ

により、故障したディスク０のＤ０を再現できる。

【０００７】また、ＲＡＩＤをはじめとするディスクアレイの制御方法には、「ハードアレイ」方式、「ソフトアレイ」方式と言われる方法がある。「ハードアレイ」方式は、ディスクアレイ制御専用のコントローラを備え、そのコントローラ上のＣＰＵで、上記のようなディスクアレイのデータ分散制御を行う方式である。「ハードアレイ」は、例えば、米国特許第５，２４９，２７９号明細書に記載されている。ここで、コントローラ上のＣＰＵを、ホスト計算機のＣＰＵと区別するため、ＭＰＵ（Ｍｉｃｒｏ Ｐｒｏｃｅｓｓｉｎｇ Ｕｎｉｔ）と称することにする。ただし、ＭＰＵは、１つで構成しても複数で構成してもよいので、これらを分け隔て無く同一の呼称を用いることにする。

【０００８】「ハードアレイ」方式では、ホスト計算機は、ディスクアレイを１台の大容量の仮想ディスクとしてとらえ、仮想ディスクに対してリード／ライト等のコマンドを発行する。ディスクアレイコントローラはこのコマンドを受信し、ディスクアレイを構成する各ディスクへのコマンドに変換する。この方式は専用のハード（コントローラ）を有することから「ハードアレイ」と称されている。

【０００９】一方、「ソフトアレイ」方式は、ディスクアレイ専用のコントローラを持たず、ホスト計算機に複数台のディスクを接続し、ＣＰＵでディスクアレイのデータ分散制御を行うものである。ホスト計算機のＣＰＵのソフトウェアでディスクアレイ制御をおこなうことから「ソフトアレイ」と称されている。

【００１０】

【発明が解決しようとする課題】最初に従来の第１の問題点について説明する。従来の「ハードアレイ」方式のディスクアレイでは、ロジカルアクセスリクエストをディスクのコマンドへの変換処理からディスクのコマンド処理まで、すべてのディスクアレイ制御を、ディスクアレイコントローラのＭＰＵで行う方式である。この方法では、多数のディスクを接続し、多数のロジカルアクセスリクエストを処理しようとすると、ＭＰＵの能力でその処理リクエスト数が限定されてしまうという問題があ

る。これは、一般に、ディスクアレイコントローラに搭載するＭＰＵの能力は、ホスト計算機のＣＰＵ能力に比べ非常に低いためである。

【００１１】また、従来の「ソフトアレイ」方式のディスクアレイでは、処理能力の高いＣＰＵで全てのディスクアレイ処理を行うので、専用のＭＰＵをもつ「ハードアレイ」より高性能が期待できる。しかし、ＣＰＵでは、負荷の重いアプリケーションや，ネットワーク，グラフィック等の処理を並行して実行しなくてはならないので、負荷の重いディスクアレイ処理をＣＰＵで実行することにより、ＣＰＵ負荷率が上昇し、本来のアプリケーション処理を妨げてしまうという問題がある。さらに、「ソフトアレイ」はホスト計算機のＯＳのファイルシステムや、デバイスドライバで実現するが、このため、ホスト計算機の種類やＯＳの種類に依存してしまうという問題がある。

【００１２】次に、従来の第２の問題点について説明する。また、従来の「ハードアレイ」方式は、ディスクアレイコントローラをホスト計算機に内蔵する方式とホスト計算機に外付けする方式がある。内蔵方式は、ホストＩ／Ｆに高速なホストバスを用い、また、ＳＣＳＩ等のオーバヘッドが無いので高性能であり、さらに、ホスト計算機にディスクアレイを内蔵できるので小型化に適するが、その反面、ホストバスの種類に依存し、さらにロジカルディスクアクセスリクエストを発行する手段が、ホスト計算機のＯＳに依存してしまい、接続性が低いという問題がある。

【００１３】一方、外付け方式は、ホストＩ／Ｆに、ＳＣＳＩ等のディスク接続バスを用いるので、ホスト計算機のホストバスやＯＳに依存しないため、接続性が高いが、その反面、ＳＣＳＩのオーバヘッドが大きく、転送性能も低いので性能が劣るという問題がある。このように両方式は一長一短の面があり、ホスト計算機と要求性能によりいずれかを選択する必要がある。このため、それぞれのホストＩ／Ｆを有するディスクアレイコントローラを別々に製作する必要があり、開発コストが高くなるという問題がある。

【００１４】次に、従来の第３の問題点について説明する。内蔵方式のハードアレイにおいては、ディスクアレイコントローラをホスト計算機に内蔵するので、複数台のホスト計算機が１台のディスクアレイを共用する構成を実現できないという問題がある。すなわち、複数台のホスト計算機で分散処理をすることで処理の高速化を図る「クラスタ構成」や、１台のホスト計算機が故障した際に待機するホスト計算機に処理を引き継ぐ「スタンバイ構成」を実現できないという問題がある。

【００１５】本発明の第１の目的は、上記第１の問題点を解決して、ディスクアレイコントローラに低価格なＭＰＵを使用しても、ＭＰＵの能力で性能が制限されない高性能なディスクアレイ制御システムを提供することにある。

【００１６】本発明の第２の目的は、上記第２の問題点を解決して、ホストバスに接続してホスト計算機に内蔵する「内蔵型ディスクアレイ」と、ＳＣＳＩ等のディスク接続バスでホスト計算機に外付け接続する「外付け型ディスクアレイ」を、唯一つのディスクアレイコントローラで実現することができるディスクアレイ制御システムを提供することにある。

【００１７】さらに、本発明の第３の目的は、ホスト計算機のホストバスやＯＳや、動作させるアプリケーションの要求性能に応じて、「ハードアレイ」方式のディスクアレイと、本発明の第１の目的を達成する高性能なディスクアレイを、選択的に切り換えることができるディスクアレイ制御システムを提供することにある。

【００１８】本発明の第４の目的は、上記第３の問題点を解決して、内蔵型ディスクアレイにおいても、クラスタ構成やスタンバイ構成等の複数台のホスト計算機で１台のディスクアレイを共用できるディスクアレイ制御システムを提供することにある。

【００１９】本発明の第５の目的は、ホスト計算機にディスクアレイを内蔵したときに、そのホスト計算機自体を一つのディスクアレイ装置として構成し得るディスクアレイ制御システムを提供することにある。

【００２０】

【課題を解決するための手段】上記第１の目的を実現するために、本発明は、プログラムを実行するＣＰＵを有するホスト計算機と、このホスト計算機にホストインターフェースを介して接続されるとともに、複数のディスクからなるディスクアレイと、このディスクアレイを制御するディスクアレイコントローラとを有するディスクアレイ装置から構成され、上記ＣＰＵは、上記ディスクアレイ装置への論理コマンドから上記ディスクアレイを構成するディスクへのディスクコマンドを生成し、かつ、冗長データを生成するための冗長データ生成コマンドを生成する上位ディスクアレイ制御手段を備え、上記ディスクアレイコントローラは、上記ディスクアレイコントローラの内部の制御プログラムを実行する上記ＣＰＵよりも処理能力の小さいＭＰＵと、上記ディスクアレイを構成するディスクを接続するディスクインターフェースと、上記ディスクアレイに格納する冗長データを生成する冗長データ生成手段とを備え、上記ＭＰＵは、上記上位ディスクアレイ制御手段が発行した上記ディスクコマンド及び冗長データ生成コマンドを受信し、上記ディスクアレイを構成するディスクにディスクコマンドを発行し、また、上記冗長データ生成手段に冗長データ生成コマンドを発行する下位ディスクアレイ制御手段を備えるようにしたものであり、かかる構成により、高性能化が図り得るものとなる。

【００２１】上記第２の目的を達成するために、本発明は、プログラムを実行するＣＰＵを有するホスト計算機

と、このホスト計算機にホストインターフェースを介して接続されるとともに、複数のディスクからなるディスクアレイと、このディスクアレイを制御するディスクアレイコントローラとを有するディスクアレイ装置から構成され、上記ディスクアレイ装置の上記ホストインターフェースは、上記ホスト計算機の内部システムバスに直結するカードエッジを有する構造の第１ホストインターフェースと、上記ホスト計算機とケーブルを介して接続する構造の第２ホストインターフェースとから構成され、上記ディスクアレイ装置は、上記第１ホストインターフェースと上記第２ホストインターフェースのいずれか一方を選択するホストインターフェースモード選択手段を備えるようにしたものであり、かかる構成により、内蔵型と外付け型を選択し得るものとなる。

【００２２】また、上記ディスクアレイ制御システムにおいて、好ましくは、上記ＣＰＵは、さらに、上記第１ホストインターフェースに上記ＣＰＵが生成したコマンドを送信し、上記ディスクアレイコントローラからの終了報告を受信する第１コマンド送信制御手段と、上記第２ホストインターフェースに上記ＣＰＵが生成したコマンドを送信し、ディスクアレイコントローラからの終了報告を受信する第２コマンド送信制御手段と、上記第１若しくは第２コマンド送信制御手段を選択するコマンド送信選択手段を備え、上記ディスクアレイコントローラは、さらに、上記第１ホストインターフェースから上記ＣＰＵが生成したコマンドを受信し、その終了を上記ホスト計算機に報告する第１コマンド受信制御手段と、上記第２ホストインターフェースから上記ＣＰＵが生成したコマンドを受信し、その終了をホストに報告する第２コマンド受信制御手段と、上記ホストインターフェースモード選択手段により決定した上記第１若しくは第２ホストインターフェースに応じた上記第１若しくは第２コマンド受信制御手段を選択するコマンド受信選択手段とを備えるようにしたものである。

【００２３】上記ディスクアレイ制御システムにおいて、好ましくは、上記ホストインターフェースモード選択手段が、上記第２ホストインターフェースを選択した際には、上記２ホストインターフェースと上記ホスト計算機はケーブルで接続し、上記ホスト計算機の外部に設けたディスクアレイ筐体に上記ディスクアレイコントローラ及び上記ディスクアレイを搭載し、上記ディスクアレイ筐体は、上記第１ホストインターフェースのカードエッジを接続する構造のコネクタと、電源と、クロック信号生成手段を備え、上記ディスクアレイコントローラは、第１ホストインターフェースのカードエッジで上記コネクタに接続し、上記コネクタを介して上記電源から電力を供給し、上記クロック信号生成手段からクロック信号を供給するようにしたものである。

【００２４】上記ディスクアレイ制御システムにおいて、好ましくは、上記ディスクアレイ筐体は、上記ホストインターフェースモード選択手段と接続するコネクタを有し、上記ホストインターフェースモード選択手段は、上記ディスクアレイ筐体のコネクタと接続した際には第２ホストインターフェース制御手段を選択するようにしたものである。

【００２５】上記第３の目的を達成するために、本発明は、プログラムを実行するＣＰＵを有するホスト計算機と、このホスト計算機にホストインターフェースを介して接続されるとともに、複数のディスクからなるディスクアレイと、このディスクアレイを制御するディスクアレイコントローラとを有するディスクアレイ装置から構成され、上記ＣＰＵ若しくは上記ＭＰＵは、上記ディスクアレイ装置への論理コマンドから上記ディスクアレイを構成するディスクへのディスクコマンドを生成し、かつ、冗長データを生成するための冗長データ生成コマンドを生成する上位ディスクアレイ制御手段を備え、上記ディスクアレイコントローラは、上記ディスクアレイコントローラの内部の制御プログラムを実行する上記ＣＰＵよりも処理能力の小さいＭＰＵと、上記ディスクアレイを構成するディスクを接続するディスクインターフェースと、上記ディスクアレイに格納する冗長データを生成する冗長データ生成手段とを備え、上記ＭＰＵは、上記上位ディスクアレイ制御手段が発行した上記ディスクコマンド及び冗長データ生成コマンドを受信し、上記ディスクアレイを構成するディスクにディスクコマンドを発行し、また、上記冗長データ生成手段に冗長データ生成コマンドを発行する下位ディスクアレイ制御手段を備え、さらに、上記ディスクアレイコントローラは、上記ＣＰＵ若しくは上記ＭＰＵが備える上記上位ディスクアレイ制御手段を選択するディスクアレイモード選択手段を備えるようにしたものであり、かかる構成により、ハイブリッドアレイ構成とハードアレイ構成を選択し得るものとなる。

【００２６】上記第４の目的を達成するために、本発明は、プログラムを実行するＣＰＵを有する複数のホスト計算機と、これらのホスト計算機にホストインターフェースを介して接続されるとともに、複数のディスクからなるディスクアレイと、このディスクアレイを制御するディスクアレイコントローラとを有するディスクアレイ装置から構成され、上記複数のホスト計算機のそれぞれの上記ＣＰＵは、上記ディスクアレイ装置への論理コマンドから上記ディスクアレイを構成するディスクへのディスクコマンドを生成し、かつ、冗長データを生成するための冗長データ生成コマンドを生成する上位ディスクアレイ制御手段を備え、上記ディスクアレイコントローラは、上記ディスクアレイコントローラの内部の制御プログラムを実行する上記ＣＰＵよりも処理能力の小さいＭＰＵと、上記ディスクアレイを構成するディスクを接続するディスクインターフェースと、上記ディスクアレイに格納する冗長データを生成する冗長データ生成手段

とを備え、上記ＭＰＵは、上記上位ディスクアレイ制御手段が発行した上記ディスクコマンド及び冗長データ生成コマンドを受信し、上記ディスクアレイを構成するディスクにディスクコマンドを発行し、また、上記冗長データ生成手段に冗長データ生成コマンドを発行する下位ディスクアレイ制御手段を備え、上記ディスクアレイ装置は、第１の上記ホスト計算機に内蔵され、上記ディスクアレイ装置の上記ホストインターフェースは、上記第１のホスト計算機の内部システムバスに直結するカードエッジを有する構造の第１ホストインターフェースと、上記第１のホスト計算機以外の上記ホスト計算機とケーブルを介して接続する構造の第２ホストインターフェースとから構成するようにしたものである。

【００２７】上記第５の目的を達成するために、本発明は、プログラムを実行するＣＰＵを有する第１のホスト計算機と、プログラムを実行するＣＰＵを有する第２のホスト計算機と、上記第１及び第２のホスト計算機にホストインターフェースを介して接続されるとともに、複数のディスクにからなるディスクアレイと、このディスクアレイを制御するディスクアレイコントローラとを有するディスクアレイ装置から構成され、上記第１の計算機の上記ＣＰＵは、上記ディスクアレイ装置への論理コマンドから上記ディスクアレイを構成するディスクへのディスクコマンドを生成し、かつ、冗長データを生成するための冗長データ生成コマンドを生成する上位ディスクアレイ制御手段を備え、上記第２の計算機の上記ＣＰＵは、上記ディスクアレイ装置への論理コマンドを上記ディスクアレイコントローラに発行するディスク制御手段を備え、上記ディスクアレイコントローラは、上記ディスクアレイコントローラの内部の制御プログラムを実行する上記ＣＰＵよりも処理能力の小さいＭＰＵと、上記ディスクアレイを構成するディスクを接続するディスクインターフェースと、上記ディスクアレイに格納する冗長データを生成する冗長データ生成手段とを備え、上記ＭＰＵは、上記上位ディスクアレイ制御手段が発行した上記ディスクコマンド及び冗長データ生成コマンドを受信し、上記ディスクアレイを構成するディスクにディスクコマンドを発行し、また、上記冗長データ生成手段に冗長データ生成コマンドを発行する下位ディスクアレイ制御手段を備え、上記ディスクアレイ装置は、第１の上記ホスト計算機に内蔵され、上記ディスクアレイ装置の上記ホストインターフェースは、上記第１のホスト計算機の内部システムバスに直結するカードエッジを有する構造の第１ホストインターフェースと、上記第２のホスト計算機とケーブルを介して接続する構造の第２ホストインターフェースとから構成するようにしたものである。

【００２８】また、本発明は、ホスト計算機にホストインターフェースを介して接続されるとともに、複数のディスクからなるディスクアレイを制御するディスクアレイコントローラにおいて、上記ディスクアレイコントローラの内部の制御プログラムを実行するＭＰＵと、上記ディスクアレイを構成するディスクを接続するディスクインターフェースと、上記ディスクアレイに格納する冗長データを生成する冗長データ生成手段とを備え、上記ホストインターフェースは、上記ホスト計算機の内部システムバスに直結するカードエッジを有する構造の第１ホストインターフェースと、上記ホスト計算機とケーブルを介して接続する構造の第２ホストインターフェースとから構成するようにしたものである。

【００２９】

【発明の実施の形態】以下、図１～図１９を用いて、本発明の一実施形態によるディスクアレイ制御システムについて説明する。

【００３０】［構成の説明］最初に、図１を用いて、本発明の一実施形態による内蔵型ハイブリッドアレイ構成のディスクアレイ制御システムの全体構成について説明する。図１は、本発明の一実施形態によるディスクアレイ制御システムのハードウエアの全体構成のブロック図である。

【００３１】ホスト計算機１は、ＣＰＵ１０と、主記憶メモリ１１と、システム制御手段１２と、ＰＣＩバス１３と、ＰＣＩコネクタ１４と、ディスクアレイ装置２によって構成されている。

【００３２】システム制御手段１２は、ＣＰＵ１０とＰＣＩバス１３と主記憶メモリ１１との間のデータ転送を制御する。ＰＣＩバス１３は、ホストバス若しくはシステムバスである。ＰＣＩコネクタ１４は、ＰＣＩバス１３にＰＣＩ拡張ボードを接続する。ディスクアレイ装置２は、ＰＣＩコネクタ１４を介して、ＰＣＩバス１３に接続されている。ディスクアレイ装置２は、ホスト計算機１の中に内蔵されており、この構成を、「内蔵型」と称する。

【００３３】ディスクアレイ装置２は、少なくとも１台以上のディスク２１ａ，２１ｂ，２１ｃ，２１ｄ，２１ｅ，２１ｆ，…と、ディスクアレイコントローラ３によって構成されている。ディスクアレイコントローラ３は、複数台のディスク２１ａ，２１ｂ，…を制御して、これらのディスクをディスクアレイとして機能させるディスクアレイコントローラ３は、ＭＰＵ３０と、メモリ４２と、ＰＣＩ　Ｉ／Ｆコントローラ３２１と、ＳＣＳＩ３３，３４ａ，３４ｂと、パリティ生成回路３５と、キャッシュメモリ３６と、ホストＩ／Ｆモード選択手段３７と、ＰＣＩバス３８とから構成されている。

【００３４】ＭＰＵ３０は、ディスクアレイコントローラ３の制御プログラムを実行するものである。メモリ４２は、ＭＰＵ３０が実行するプログラムやデータを格納している。

【００３５】ＰＣＩ　Ｉ／Ｆコントローラ３２１は、ホスト計算機１の第１ホストＩ／ＦであるＰＣＩＩ／Ｆ３

２を制御する。ＰＣＩＩ／Ｆ３２は、ＰＣＩコネクタ１４とＰＣＩ　Ｉ／Ｆコントローラ３２１によって構成される。また、ＰＣＩ　Ｉ／Ｆコントローラ３２１は、ＤＭＡコントローラ（図示せず）を内蔵している。

【００３６】ＳＣＳＩ３３は、ホスト計算機１と接続するための第２ホストＩ／Ｆである。ＳＣＳＩ３３は、ＰＣＩカードエッジ３３２と、ＳＣＳＩコントローラ３３１によって構成されている。ＳＣＳＩコントローラ３３１は、ＳＣＳＩ３３を制御する。ＳＣＳＩコントローラ３３１は、ＤＭＡコントローラ（図示せず）を内蔵している。ＰＣＩカードエッジ３３２は、ＳＣＳＩケーブルを接続するＳＣＳＩコネクタであるが、本実施形態においては、ＳＣＳＩケーブルは接続されていない。

【００３７】ＳＣＳＩ３４ａは、少なくとも１台以上のディスク２１ａ，２１ｂ，２１ｃ，…を接続するためのディスクＩ／Ｆである。ＳＣＳＩ３４ａは、ＳＣＳＩコントローラ３４ａ１と、ＳＣＳＩコネクタ４０ａ１によって構成されている。ＳＣＳＩコントローラ３４ａ１は、ＳＣＳＩ３４ａを制御する。ＳＣＳＩコントローラ３４ａ１は、ＤＭＡコントローラ（図示せず）を内蔵している。ＳＣＳＩコネクタ４０ａ１は、ＳＣＳＩケーブルを接続するコネクタである。

【００３８】ＳＣＳＩ３４ｂは、少なくとも１台以上のディスク２１ｄ，２１ｅ，２１ｆ，…を接続するためのディスクＩ／Ｆである。ＳＣＳＩ３４ｂは、ＳＣＳＩコントローラ３４ｂ１と、ＳＣＳＩコネクタ４０ｂ１によって構成されている。ＳＣＳＩコントローラ３４ｂ１は、ＳＣＳＩ３４ｂを制御する。ＳＣＳＩコントローラ３４ｂ１は、ＤＭＡコントローラ（図示せず）を内蔵している。ＳＣＳＩコネクタ４０ｂ１は、ＳＣＳＩケーブルを接続するコネクタである。

【００３９】パリティ生成回路３５は、ディスクアレイに格納する冗長データであるパリティデータを生成する。キャッシュメモリ３６は、ディスク２１ａ，２１ｂ，２１ｃ，２１ｄ，２１ｅ，２１ｆ，…のデータを一時的に保持する。

【００４０】ホストＩ／Ｆモード選択手段３７は、ホスト計算機１とディスクアレイコントローラ３の接続を上記第１ホストＩ／ＦであるＰＣＩ　Ｉ／Ｆ３２を用いるか、第２ホストＩ／ＦであるＳＣＳＩ３３を用いるかを選択するものである。

【００４１】ＰＣＩバス３８は、ディスクアレイコントローラ３の内部のバスである。

【００４２】同図において、ホスト計算機１のホストバス及びディスクアレイコントローラ３の内部バスにＰＣＩバス１３，３８を用いているが、他のバスを用いてもよい。

【００４３】また、ディスクＩ／Ｆ３４ａ，３４ｂにＳＣＳＩを用いているがこれもこの限りではない。また、ＳＣＳＩコントローラ３４の数を２としているが、これはそれ以上でも以下でもよい。また、ＰＣＩ　Ｉ／Ｆコントローラ３２１とＭＰＵ３０は別の構成にしているが、ＭＰＵ３０がＰＣＩ　Ｉ／Ｆを備えているものであってもよい。また、キャッシュメモリ３６の制御回路は、キャッシュメモリ３６に含まれているものとし、パリティ生成回路３５は独立した構成にしているが、パリティ生成回路をキャッシュメモリ制御回路と一体化して、キャッシュメモリ３６内に含まれるものとしてもよい。このように、本実施形態は、様々な構成のバリエーションに適用できるものであり、本実施形態は本発明の範囲を限定するものではない。

【００４４】次に、図２を用いて、本発明の一実施形態による内蔵型ハイブリッドアレイ構成のディスクアレイ制御システムのＣＰＵ１０及びＭＰＵ３０で実行するプログラムの構成について説明する。図２は、本発明の一実施形態によるディスクアレイ制御システムのソフトウエアの全体構成のブロック図である。

【００４５】本実施形態においては、ディスクアレイ処理を、ホスト計算機１のＣＰＵ１０とディスクコントローラ３のＭＰＵ３０とで分散しており、この処理構成を「ハイブリッドアレイ」構成と称する。

【００４６】ＣＰＵ１０が実行するプログラムは、アプリケーション１００と、ＯＳ１０１と、ファイルシステム１０２と、上位ディスクアレイ制御ドライバ１０３と、ＰＣＩ　Ｉ／Ｆドライバ１０４と、ＳＣＳＩドライバ１０５から構成されている。

【００４７】ここで、上位ディスクアレイ制御ドライバ１０３は、論理コマンドアドレス変換部１０３１と、キャッシュ管理部１０３２と、コマンドＩ／Ｆ制御部１０３４と、コントローラ検出部１０３５とから構成されている。論理コマンドアドレス変換部１０３１は、ファイルシステム１０２から受信したディスクアレイへの論理コマンドをディスクアレイ装置２を構成する少なくとも１台以上のディスク２１へのコマンドに変換するものである。キャッシュ管理部１０３２は、キャッシュメモリ３６の管理を行うものである。コマンドＩ／Ｆ制御部１０３４は、ＰＣＩ　Ｉ／Ｆドライバ１０４およびＳＣＳＩドライバ１０５のいずれか選択されたドライバとの間でコマンド送信等のやりとりを行うものである。コントローラ検出部１０３５は、ＰＣＩ　Ｉ／Ｆ３２とＳＣＳＩ３３のどちらのホストＩ／Ｆを使用するかを選択し、ディスクアレイコントローラの検出，初期化を行うものである。

【００４８】ＰＣＩ　Ｉ／Ｆドライバ１０４は、ＰＣＩ　Ｉ／Ｆ３２を用いてディスクアレイコントローラ３にコマンドを送信するためのものであり、第１コマンド送信手段である。ＳＣＳＩドライバ１０５は、ＳＣＳＩ３３を用いてディスクアレイコントローラ３にコマンドを送信するためのものであり、第２コマンド送信手段である。

【００４９】また、ＭＰＵ３０が実行するプログラムは、ＰＣＩ　Ｉ／Ｆコマンド制御部３０１と、ＳＣＳＩコマンド制御部３０２と、下位ディスクアレイ制御部３０３と、ホストＩ／Ｆモード制御部３０５とから構成されている。ＰＣＩ　Ｉ／Ｆコマンド制御部３０１は、ＰＣＩ　Ｉ／Ｆ３２を介してホスト計算機１からのコマンドを受信するものであり、第１コマンド受信手段である。ＰＣＩ　Ｉ／Ｆコマンド制御部３０１は、ＤＭＡ制御部を含んでおり、このＤＭＡ制御部は、ホスト計算機１の主記憶メモリ１１とディスクアレイコントローラ３のキャッシュメモリ３６間のデータ転送制御を実行する。ＳＣＳＩコマンド制御部３０２は、ＳＣＳＩ３３を介してホスト計算機１からのコマンドを受信するものであり、第２コマンド受信手段である。ＳＣＳＩコマンド制御部３０２は、ＤＭＡ制御部を含んでおり、このＤＭＡ制御部は、ホスト計算機１の主記憶メモリ１１とディスクアレイコントローラ３のキャッシュメモリ３６間のデータ転送制御を実行する。なお、図１に示す内蔵型においては、ＳＣＳＩコマンド制御部３０２は、使用されていない。下位ディスクアレイ制御部３０３は、上位ディスクアレイ制御ドライバ１０３の生成したディスクコマンドやパリティ生成コマンドを実行するものである。ホストＩ／Ｆモード制御部３０５は、ホストＩ／Ｆモード選択手段３７により選択されたホストＩ／Ｆで、ホスト計算機１とコマンドの授受をできるように制御する。図１に示す構成においては、ホストＩ／Ｆモード制御部３０５からの制御信号によって、ＰＣＩ　Ｉ／Ｆ３２がホストＩ／Ｆとして選択されており、ホスト計算機１とコマンドの授受をできるように制御する。

【００５０】下位ディスクアレイ制御部３０３は、コマンドＩ／Ｆ制御部３０３１と、ディスクコマンド制御部３０３２と、パリティ生成コマンド制御部３０３３と、初期化部３０３５と、コマンド管理部３０３６とから構成されている。コマンドＩ／Ｆ制御部３０３１は、ホストＩ／Ｆモード制御部３０５により決定されたＰＣＩ　Ｉ／Ｆ制御部３０１もしくはＳＣＳＩ制御部３０２のどちらか一方からコマンドを受信するものである。ディスクコマンド制御部３０３２は、ディスクＩ／Ｆ３４ａ，３４ｂのＳＣＳＩコントローラ３４ａ１，３４ｂ１を制御し、ディスクコマンドを実行するものである。パリティ生成コマンド制御部３０３３は、パリティ生成回路３５を制御し、パリティ生成コマンドを実行するものである。初期化部３０３５は、ディスクアレイコントローラ３の初期化処理を実行するものである。コマンド管理部３０３６は、受信したコマンドの管理を行うものである。

【００５１】［動作の説明］

（１）ホストＩ／Ｆモードの選択

次に、図３及び図４を用いて、本発明の一実施形態による内蔵型ハイブリッドアレイ構成のディスクアレイ制御システムにおけるホストＩ／Ｆモード選択手段３７の構成及び機能について説明する。図３は、本発明の一実施形態によるディスクアレイ制御システムのホストＩ／Ｆモード選択手段の構成を示す回路図であり、図４は、本発明の一実施形態によるディスクアレイ制御システムのホストＩ／Ｆモード選択手段の論理図である。

【００５２】ホストＩ／Ｆモード選択手段３７は、例えば，図３に示すような構成を有している。スイッチ３７１の開閉により、ホストＩ／Ｆモード選択手段３７の出力信号であるＩＦ＿Ｍｏｄｅ信号３７２は、”１”と”０”の２値を持つことができる。

【００５３】図４に示すように、ホストＩ／Ｆモード制御部３０５は、ＩＦ＿Ｍｏｄｅ信号３７２に応じて、ホストＩ／Ｆを選択する。即ち、ＩＦ＿Ｍｏｄｅ信号３７２が”１”の時、ホストＩ／Ｆモード制御部３０５は、ホストＩ／Ｆとして、第１ホストＩ／ＦであるＰＣＩ　Ｉ／Ｆ３２を選択し、ＩＦ＿Ｍｏｄｅ信号が”０”の時、ホストＩ／Ｆとして、第２ホストＩ／ＦであるＳＣＳＩ３３を選択する。図３に示す状態では、ＩＦ＿Ｍｏｄｅ信号３７２は”１”であるので、ＰＣＩＩ／Ｆ３２を選択していることを示している。

【００５４】以下の説明においては、ホストＩ／Ｆモード制御部３０５は、ＩＦ＿Ｍｏｄｅ信号３７２に基づいてＰＣＩ　Ｉ／Ｆ３２を選択し、図１に示したように、ディスクアレイコントローラ３がホスト計算機１のＰＣＩコネクタ１４に接続している場合について説明する。

【００５５】ディスクアレイ装置２の電源投入時に、ＭＰＵ３０は初期化を開始する。初期化開始後、ＭＰＵ３０は、メモリ４２からプログラムをリードし、下位ディスクアレイ制御部３０３が動作可能なように初期化を実施する。このとき、ホストＩ／Ｆモード選択手段３７が出力するＩＦ＿Ｍｏｄｅ３７２信号は、ＭＰＵ３０に入力され、下位ディスクアレイ制御部３０３のホストＩ／Ｆモード制御部３０５は、選択されたホストＩ／Ｆに対応するコマンド受信手段を選択する。

【００５６】図３に示す場合においては、ＰＣＩ　Ｉ／Ｆ３２が選択されているので、ＭＰＵ３０は、ＰＣＩ　Ｉ／Ｆ制御部３０１を以降使用する。また、コマンドＩ／Ｆ制御部３０３１をＰＣＩ　Ｉ／Ｆ制御部３０１と接続するよう初期化する。さらに、ＳＣＳＩコントローラ３３１，３４ａ１，３４ｂ１やパリティ生成回路３５等のハードウェアの初期化も実施する。

【００５７】また、以降、選択したホストＩ／Ｆとは異なるＩ／Ｆからコマンドが送信されてもディスクアレイコントローラ３は応答しないか、エラー応答を返すか何れかの方法でホスト計算機に以後のコマンド発行を抑止するよう知らせる。

【００５８】（２）コントローラの検出

次に、図５を用いて、本発明の一実施形態による内蔵型ハイブリッドアレイ構成のディスクアレイ制御システム

におけるディスクアレイコントローラの検出及び初期化処理について説明する。図５は、本発明の一実施形態によるディスクアレイ制御システムにおけるディスクアレイコントローラ検出処理を説明するフローチャートである。

【００５９】図１に示すホスト計算機１に電源が投入されると、図２に示したＯＳ１０１は、初期化処理の段階で、各周辺機器用のデバイスドライバを組み込み、デバイスドライバは周辺機器の検出、初期化を行う。ディスクアレイ装置２についても同様である。

【００６０】ＯＳ１０１は、上位ディスクアレイ制御ドライバ１０３を組み込む。上位ディスクアレイ制御ドライバ１０３のコントローラ検出部１０３５は、図５に示すフローチャートに従い、ディスクアレイコントローラの検，、初期化処理を行う。

【００６１】図５のステップ１０００において、上位ディスクアレイ制御ドライバ１０３がロードされ、処理が開始される。最初に、ステップ１００１において、上位ディスクアレイ制御ドライバ１０３のコントローラ検出部１０３５は、ＰＣＩ　Ｉ／Ｆ３２を選択する。次に、ステップ１００２において、コントローラ検出部１０３５は、ＰＣＩバス１３を走査し、ディスクアレイコントローラ３があるかどうか検査する。具体的には、ＰＣＩバス１３に接続する全てのデバイス（コントローラ）のベンダＩＤとデバイスＩＤを示すレジスタをリードし、ディスクアレイコントローラ３のＰＣＩ　Ｉ／Ｆコントローラ３２１の持つベンダＩＤ，デバイスＩＤを検査する。

【００６２】ステップ１００３において、検出の有無を判断し、検出した場合には、ステップ１００８において、コントローラ検出部１０３５は、ディスクアレイコントローラ３へのコマンド発行等の通信には、ＰＣＩ　Ｉ／Ｆ３２を使用することとし、上位ディスクアレイ制御ドライバ１０３とＰＣＩ　Ｉ／Ｆドライバ１０４とを接続するようコマンドＩ／Ｆ制御部１０３４を初期化する。また、ＰＣＩバス１３上にディスクアレイコントローラ３を検出できなかったときには、ステップ１００４において、コントローラ検出部１０３５は、ＳＣＳＩを選択する。次に、ステップ１００５において、コントローラ検出部１０３５は、ＳＣＳＩを走査し、ディスクアレイコントローラ３があるかどうか検査する。具体的には、ＳＣＳＩに接続する全てのデバイス（コントローラ）にコントローラ種別を問いかけるコマンド（Ｉｎｑｕｉｒｙコマンド）をＳＣＳＩドライバ経由１０５で発行し、その返値から、ディスクアレイコントローラ３が存在するかどうか特定する。

【００６３】ステップ１００６において、検出の有無を判断し、検出した場合には、以降、ディスクアレイコントローラ３へのコマンド発行等の通信にはＳＣＳＩを使用することとし、上位ディスクアレイ制御ドライバ１０３とＳＣＳＩドライバ１０５とを接続するようコマンドＩ／Ｆ制御部１０３４を初期化する。

【００６４】なお、ステップ１００４，１００５，１００６における処理は、ＰＣＩバス１３にＳＣＳＩが接続され、このＳＣＳＩとディスクアレイコントローラ３のＳＣＳＩ３３が接続される場合の処理であり、このような接続関係を「外付け型」と称し、「外付け型」の詳細については、図２０を用いて後述する。

【００６５】ステップ１０１０において、上位ディスクアレイ制御ドライバ１０３のコントローラ検出部１０３５は、上述したように、何れかのホストＩ／Ｆ上にディスクアレイコントローラ３を検出したならば、ディスクアレイコントローラの初期化コマンドを生成し、各ホストＩ／Ｆに対応するＰＣＩ　Ｉ／Ｆドライバ１０４，若しくはＳＣＳＩドライバ１０５を介して、ディスクアレイコントローラ３にディスクアレイ構成情報リードコマンドを発行する。ディスクアレイコントローラ３のＰＣＩ　Ｉ／Ｆ制御部３０１若しくはＳＣＳＩ制御部３０２は、ディスクアレイ構成情報リードコマンドを受信し、下位ディスクアレイ制御部３０３の初期化部３０３５は、ディスク２１の特定の領域に格納しているディスクアレイの構成情報をリードし、ホスト計算機１にディスクアレイ構成情報を転送する。

【００６６】次に、ステップ１０１１において、コントローラ検出部１０３５は、このディスクアレイ構成情報に基づき、上位ディスクアレイ制御ドライバ１０３の初期設定を実施する。

【００６７】また、もし、ステップ１００６において、コントローラ検出部１０３５が、何れのホストＩ／Ｆにもディスクアレイコントローラ３が検出できなかったときは、ステップ１００７において、上位ディスクアレイ制御ドライバはアンロードされる。

【００６８】（３）ディスクアレイの動作

次に、図１に示すように、ホストＩ／ＦとしてＰＣＩ　Ｉ／Ｆ３２を選択した時のディスクアレイの動作について説明する。

【００６９】（ａ）リード動作

次に、図６を用いて、本発明の一実施形態による内蔵型ハイブリッドアレイ構成のディスクアレイ制御システムにおけるディスクアレイのリード動作について説明する。図６は、本発明の一実施形態によるディスクアレイ制御システムにおけるディスクアレイのリード処理の説明図である。

【００７０】アプリケーション１００からディスクアレイ装置２へのリード要求が発行されたものとする。リードデータＡは、主記憶メモリ１１の領域１１１にリードされる。リードデータＡは、ディスクアレイ装置２のストライプサイズより大きく、図６に示すように、ディスク２１ａのストライプｓａに格納されたデータブロックａと、ディスク２１ｂのストライプｓｂに格納されたデ

ータブロックｂと、ディスク２１ｃのストライプｓｃの一部に格納されたデータブロックｃとによって構成されているものとする。

【００７１】次に、図７～図１５を用いて、ディスクアレイのリード動作について順次説明する。

【００７２】（ｉ）　コマンドリンクの発行

アプリケーション１００が、リードデータＡのリード要求を発行する。ファイルシステム１０２は、このリード要求を受信し、ディスクアレイ２１への論理リードコマンドを上位ディスクアレイ制御ドライバ１０３に発行する。

【００７３】次に、図７を用いて、本発明の一実施形態によるディスクアレイ制御システムにおける上位ディスクアレイ制御ドライバ１０３のコマンドリンクの発行方法について説明する。図７は、本発明の一実施形態によるディスクアレイ制御システムにおける上位ディスクアレイ制御ドライバのコマンドリンクの発行方法について説明するフローチャートである。

【００７４】図７のステップ１１００において、上位ディスクアレイ制御ドライバ１０３の論理コマンドアドレス変換部１０３１は、論理リードコマンドを受信する。ステップ１１０１において、論理コマンドアドレス変換部１０３１は、受信した論理リードコマンドの論理アドレスを、ディスクアレイを構成する少なくとも１台以上の個々のディスクのアドレスに変換する。この時点で、図６に示すように、リードデータＡがディスク２１ａ，２１ｂ，２１ｃのストライプｓａ，ｓｂ，ｓｃに格納するデータブロックａ，ｂ，ｃにより構成されていることがわかる。

【００７５】次に、ステップ１１０２において、キャッシュ管理部１０３２は、データブロックａ，ｂ，ｃ毎に、ディスクアレイコントローラ３のキャッシュメモリ３６のヒットミス判定を実施する。図６に示す例においては、データｂはヒット、データａ，ｃはミスヒットであったとする。ステップ１１０３において、ヒット，ミスヒットの判定を行い、ヒット時には、ステップ１１０４において、キャッシュ管理部１０３２は、キャッシュメモリ３６から主記憶メモリ１１にデータを転送するためのディスクアレイコントローラへのコマンドである「キャッシュリードコマンド」を生成する。以下、ディスクアレイコントローラへのコマンドを「コントローラコマンド」と称する。なお、図７のステップ１１０４に記載した②「キャッシュライトコマンド」は、ライト動作時のコマンドであり、これについては、ライト動作の説明において後述する。

【００７６】ミスヒット時には、ステップ１１０５において、キャッシュ管理部１０３２は、キャッシュメモリ３６にデータをロードする領域を割り当てる。そして、ステップ１１０６において、キャッシュ管理部１０３２は、ディスク２１からキャッシュメモリ３６にロード後、主記憶メモリ１１にデータを転送するための「キャッシュロード＆リードコマンド」を生成する。なお、図７のステップ１１０６に記載した②「キャッシュロード＆ライトコマンド」及び③「キャッシュライトコマンド」は、ライト動作時のコマンドであり、これについては、ライト動作の説明において後述する。

【００７７】ステップ１１０７において、キャッシュ管理部１０３２は、生成したコントローラコマンドについて、同一論理コマンド毎にコマンドリンクを生成し、主記憶メモリの特定の領域に格納する。ステップ１１０８において、キャッシュ管理部１０３２は、ステップ１１０２～１１０７の処理を、論理コマンドの扱うリードデータを構成する全てのデータブロックについて実施する。

【００７８】図６に示す例では、データブロックａ，ｃに対応するキャッシュロード＆リードコマンドと、データブロックｂに対応するキャッシュリードコマンドの３つのコマンドがリンクされ、主記憶メモリ１１に格納される。

【００７９】以下、このコントローラコマンドをリンクしたコマンド群を「コマンドリンク」と称する。この例の場合、データブロックｃは、ストライプｃの一部データのみをリードするので、キャッシュロード＆リードコマンドには、ディスク２１ｃ上のストライプｃの先頭アドレスと、データブロックｃの先頭アドレスの両者を格納している。

【００８０】次に、ステップ１１０９において、コマンドＩ／Ｆ制御部１０３は、ＰＣＩ／Ｆドライバ１０４に指示し、コマンドリンクをディスクアレイコントローラ３に発行する。

【００８１】ここで、図８を用いて、ホスト計算機１とディスクアレイコントローラ３の間のコマンドリンクの発行の方法について説明する。図８は、本発明の一実施形態によるディスクアレイ制御システムにおけるホスト計算機とディスクアレイコントローラの間のコマンドリンクの発行方法の説明図である。

【００８２】ディスクアレイコントローラ３のＰＣＩ　Ｉ／Ｆコントローラ３２１は、図８に示すように、ＤＭＡコントローラ３２１１と、ＭａｉｌＢｏｘレジスタ３２１２とＤｏｏｒＢｅｌｌレジスタ３２１３の2つのレジスタを備えている。ＣＰＵ１０のＰＣＩ　Ｉ／Ｆドライバ１０４は、主記憶メモリ１１上のコマンドリンク１１のアドレスとその識別番号を含む「コマンドリンク発行パケット」を、ＭａｉｌＢｏｘレジスタ３２１２にライトする（図８（イ））。すると、ＰＣＩ／Ｆコントローラ３２１は、ＭＰＵ３０に割り込みを発行する（図８（ロ））。

【００８３】次に、図９を用いて、ＰＣＩ　Ｉ／Ｆコントローラ３２１から割り込みを受信したＭＰＵ３０の処理について説明する。図９は、本発明の一実施形態によ

るディスクアレイ制御システムにおけるＭＰＵのコマンドリンクの割り込み処理について説明するフローチャートである。

【００８４】図９のステップ１２００において、ＭＰＵ３０は割り込みを受信する。

【００８５】次に、ステップ１２０１において、ＭＰＵ３０のＰＣＩ　Ｉ／Ｆ制御部３０１は、ＭａｉｌＢｏｘレジスタ３２１２をリードし、「コマンドリンク発行パケット」が送信されたことを認識し、ＤｏｏｒＢｅｌｌレジスタ３２１３の「パケット受領ビット」をセットする。ステップ１２０２において、ＰＣＩ　Ｉ／Ｆ制御部３０１は、コマンドリンクをホスト計算機１の主記憶メモリ１１からディスクアレイコントローラ３のメモリ４２に転送するための「ＤＭＡコマンド」を生成する。さらに、ステップ１２０３において、ＰＣＩ　Ｉ／Ｆ制御部３０１は、ＰＣＩ　Ｉ／Ｆ制御部３０１が備えているＤＭＡ制御部のＤＭＡコマンドキューにキィーイングする。ここで、図７のステップ１１１０に戻り、ホスト計算機１のＰＣＩ　Ｉ／Ｆドライバ１０４は、ＤｏｏｒＢｅｌｌレジスタ３２１３の「パケット受領ビット」がセットされたことを確認する。

【００８６】以上の処理によって、コマンドリンクの発行処理を終了する。以降、コマンドリンクがすべて実行終了し、終了割り込みを受信するまで、ＣＰＵ１０はその他の処理を実行することができる。

【００８７】次に、図１０～図１４を用いて、ディスクアレイコントローラ３の内部処理について説明する。

【００８８】（ｉｉ）コマンドリンクの受信

次に、図６及び図１０を用いて、ＤＭＡコマンドキューにコマンドがキューイングされた時の、ディスクアレイコントローラ３のＭＰＵ３０のＰＣＩ　Ｉ／Ｆ制御部３０１の中のＤＭＡ制御部の処理について説明する。図１０は、本発明の一実施形態によるディスクアレイ制御システムにおけるＭＰＵのＰＣＩ　Ｉ／Ｆ制御部の中のＤＭＡ制御部の処理について説明するフローチャートである。

【００８９】ステップ１２１０において、ＰＣＩ　Ｉ／Ｆ制御部３０１の中のＤＭＡ制御部は、ＤＭＡコマンドキューの先頭のＤＭＡコマンドを取り出す。次に、ステップ１２１１において、ＰＣＩ　Ｉ／Ｆ制御部３０１の中のＤＭＡ制御部は、ＰＣＩ　Ｉ／Ｆコントローラ３２１のＤＭＡコントローラ３２１１にＤＭＡコマンドを発行し、ＤＭＡを起動する（図８（ハ））。ＤＭＡが起動すると、コマンドリンク１１１が、ホスト計算機１の主記憶メモリ１１からディスクコントローラ３のメモリ４２にＤＭＡ転送される（図８（ニ））。ＰＣＩ　Ｉ／Ｆ制御部３０１の中のＤＭＡ制御部は、ＤＭＡ転送が終了するまでの間スリープし、ＭＰＵ３０は他の処理を実行する事ができる。

【００９０】次に、ＤＭＡ転送が終了すると、ＰＣＩ　Ｉ／Ｆコントローラ３２１のＤＭＡコントローラ３２１１は、割り込みを発行し（図８（ホ））、ステップ１２１２において、この割り込みをＭＰＵ３０は受信し、ＰＣＩ　Ｉ／Ｆ制御部３０１の中のＤＭＡ制御部は動作を再開する。次に、ステップ１２１３の中のステップ１２１３ａにおいて、ＰＣＩ　Ｉ／Ｆ制御部３０１の中のＤＭＡ制御部は、下位ディスクアレイ制御部３０３のコマンドＩ／Ｆ制御部３０３１のホストコマンドキューに、受信したコマンドリンクをキューイングし、ＤＭＡコマンドキューから終了したＤＭＡコマンドを削除する。なお、ステップ１２１３ｂは、データ転送時の処理であり、これについては、後述する。

【００９１】ステップ１２１４において、ＰＣＩ　Ｉ／Ｆ制御部３０１の中のＤＭＡ制御部は、ＤＭＡコマンドキューにまだコマンドがあるか否かを判断し、ＤＭＡコマンドキューにまだコマンドがあるならば、ステップ１２１０に戻り、ステップ１２１０～１２１３を繰り返すものである。

【００９２】（ｉｉｉ）コントローラコマンド（コマンドリンク）の起動処理

次に、図１１及び図１２を用いて、コマンドＩ／Ｆ制御部３０３１のホストコマンドキューにコマンドリンクがキューイングされた時の、下位ディスクアレイ制御部３０３における処理について説明する。図１１及び図１２は、本発明の一実施形態によるディスクアレイ制御システムにおけるＭＰＵの下位ディスクアレイ制御部の処理について説明するフローチャートである。なお、図１１におけるＸ，Ｙ，Ｚは、それぞれ、図１２のＸ，Ｙ，Ｚに連結される。

【００９３】ステップ１２２０において、コマンドＩ／Ｆ制御部３０３１のホストコマンドキューにコマンドリンクがキューイングされると、下位ディスクアレイ制御部３０３のコマンド管理部３０３６は、ホストコマンドキューからコマンドリンクを一つ取り出す。

【００９４】次に、ステップ１２２１において、下位ディスクアレイ制御部３０３のコマンド管理部３０３６は、そのコマンドリンクに登録されたコントローラコマンドを一つ取り出し、ステップ１２２１において、そのコントローラコマンドを解析する。解析されたコントローラコマンドの種類によって、ステップ１２２７の「キャッシュリードコマンド」と、ステップ１２２３の「キャッシュロード＆リードコマンド」と、ステップ１２３４の「キャッシュライトコマンド」と、ステップ１２３０の「キャッシュロード＆ライトコマンド」と、ステップ１２３６の「キャッシュロード＆デステージコマンド」と、ステップ１２４２の「パリティ生成＆デステージコマンド」と、ステップ１２４６の「キャッシュデステージコマンド」との各処理に分かれる。これらの各処理の詳細については、後述する。

【００９５】ステップ１２４９において、コマンド管理

部３０３６は、全コントローラコマンドの処理が終了したか否かを判断し、全コントローラコマンドの処理が終了するまで、各コントローラコマンド毎の処理が終了すると、次のコントローラコマンドを取り出し、以下コマンドリンクに登録されたすべてのコントローラコマンドの起動処理を行う。

【００９６】全コントローラコマンドの処理が終了すると、ステップ１２５０において、コマンド管理部３０３６は、ホストコマンドキューから次のコマンドリンクを取り出し、上記の起動処理を繰り返し、すべてのキューイングされたコマンドリンクの起動が終了するとコマンド管理部３０３６は処理を終了する。

【００９７】ここで、図６に示した例では、コントローラコマンドとして、データブロックａ，ｃに対する「キャッシュロード＆リードコマンド」と、データブロックｂに対する「キャッシュリードコマンド」が使用される。データブロックａ，ｂ，ｃの順にコントローラコマンドの起動処理が行われる。以下、この２つのコマンドについて、コマンド管理部３０３６の動作を図１１を用いて説明する。

【００９８】図１１のステップ１２２３において、コマンド管理部３０３６が、「キャッシュロード＆リードコマンド」と判断すると、ステップ１２２４に進む。ステップ１２２４において、コマンド管理部３０３６は、当該ストライプの上位ディスクアレイ制御ドライバ１０３の指定したキャッシュメモリ３６のアドレスに、ディスク２１からデータをロードするためのディスクリードコマンドを生成する。なお、本実施形態では、必ずキャッシュへのロードは、ストライプ単位であるとする。よって、データブロックｃは、アプリケーションからのリード要求はストライプの一部ではあるが、上位ディスクアレイ制御ドライバの指示に基づき、ディスク２１ｃからキャッシュメモリ３６へのロードは、ストライプｓｃ全部をロードすることになる。

【００９９】ステップ１２２５において、コマンド管理部３０３６は、生成したディスクコマンドを、ディスクコマンド制御部３０３２のディスクコマンドキューにキューイングする。さらに、ステップ１２２６において、コマンド管理部３０３６は、このコントローラコマンドを「コマンドペンディング状態」に設定し、このコントローラコマンドの処理は一旦中断し、コントローラコマンドの終了を待つものである。次に、コマンド管理部３０３６は、図１２のステップ１２４９に進み、次のコントローラコマンドの処理を続行する。

【０１００】ここで、「コマンドペンディング状態」とは、「キャッシュロード＆リードコマンド」のように、（１）第１フェーズ：ディスクからキャッシュメモリへのデータのロードと、（２）第２フェーズ：キャッシュメモリから主記憶メモリへのデータのリードの２つの処理を逐次的に行う必要がある場合、第１フェーズを起動後、一旦、このコントローラコマンドを「コマンドペンディング状態」に設定し、第１フェーズが終了後、第２フェーズの処理を継続するよう制御するためのものである。「コマンドペンディング状態」のコントローラコマンドの第１フェーズが終了した後に、図１１及び図１２に点線で示すコントローラコマンド種の処理を、第２フェーズとして実行する。同図の他のコントローラコマンド処理の点線はすべて第１フェーズ終了後の第２フェーズ、若しくは、第２フェーズ終了後の第３フェーズを示している。

【０１０１】「キャッシュロード＆リードコマンド」の例においては、ディスクコマンド終了後は、図１１の点線に従い、以下に説明する「キャッシュリードコマンド」の処理を実施する。

【０１０２】図１１のステップ１２２７において、コマンド管理部３０３６が、「キャッシュリードコマンド」と判断すると、ステップ１２２８に進む。ステップ１２２８において、コマンド管理部３０３６は、キャッシュメモリ３６のデータを主記憶メモリ１１にＤＭＡ転送するためのキャッシュリードＤＭＡコマンドを生成する。そして、ステップ１２２９において、コマンド管理部３０３６は、キャッシュリードＤＭＡコマンドをＤＭＡコマンドキューにキューイングし、キャッシュリードコマンドの終了を待つものである。

【０１０３】なお、ステップ１２３４の「キャッシュライトコマンド」と、ステップ１２３０の「キャッシュロード＆ライトコマンド」と、ステップ１２３６の「キャッシュロード＆デステージコマンド」と、ステップ１２４２の「パリティ生成＆デステージコマンド」と、ステップ１２４６の「キャッシュデステージコマンド」との各処理の詳細については、ライト処理において後述する。

【０１０４】（ｉｖ）　ディスクコマンドの処理

次に、図１３を用いて、ディスクコマンドキューにコマンドがキューイングされた時の、下位ディスクアレイ制御部３０３のディスクコマンド制御部３０３２における処理について説明する。図１３は、本発明の一実施形態によるディスクアレイ制御システムにおけるＭＰＵの下位ディスクアレイ制御部のディスクコマンド制御部の処理について説明するフローチャートである。

【０１０５】ステップ１２６０において、ディスクコマンド制御部３０３２は、ディスクコマンドキューから一つのディスクコマンドを取り出す。ステップ１２６１において、ディスクコマンド制御部３０３２は、取りだしたディスクコマンドをＳＣＳＩコントローラ３４に発行して、ディスクコマンドを起動する。ステップ１２６２において、ディスクコマンド制御部３０３２は、まだディスクコマンドキューにディスクコマンドがあるか否かを判断し、ディスクコマンドキューにディスクコマンドが残っており、未動作のディスクがあったり、１台のデ

ィスクが複数のコマンドを受信できるときには、ディスクコマンドの起動を繰り返す。ディスクコマンドが全て無くなる、若しくはそれ以上の起動ができない時には、ディスクコマンド制御部３０３２は、処理を停止し、スリープする。この間、ＭＰＵ３０は、他の処理を実行できる。

【０１０６】ＳＣＳＩコントローラ３４は、内部にＤＭＡコントローラを備えており、ＤＭＡコントローラが、キャッシュメモリ３６とＳＣＳＩコントローラ３４間のデータ転送を実行する。

【０１０７】データ転送が終了すると、ＳＣＳＩコントローラ３４は終了割り込みをＭＰＵ３０に発行する。ステップ１２６３において、ＭＰＵ３０は、終了割込みを受信し、ディスクコマンド制御部３０３２は、処理を再開する。

【０１０８】ステップ１２６４において、ディスクコマンド制御部３０３２は、終了報告をコマンド管理部３０３６に対し行う。ステップ１２６５において、ディスクコマンド制御部３０３２は、終了したディスクコマンドをディスクコマンドキューから削除する。

【０１０９】ステップ１２６６において、ディスクコマンド制御部３０３２は、ディスクコマンドキューに未起動のコマンドがあるか否かを判断し、ディスクコマンドキューに未起動のコマンドがある場合には、上記の起動処理を実行し、もしなければ処理を終了する。

【０１１０】（ｖ）　ＤＭＡコマンドの処理

キャッシュメモリ３６と主記憶メモリ１１間のデータのＤＭＡ転送は、上述した（ｉ）コマンドリンクの受信と同様、ＰＣＩ　Ｉ／Ｆコントローラ３２１のＤＭＡコントローラ３２１１で実行する。

【０１１１】処理の手順は、上述した（ｉ）コマンドリンクの受信とほぼ同様であるが、相違点は、図１０のステップ１２１３ａの代わりに、ステップ１２１３ｂにおいて、データ転送終了時には、コマンド管理部３０３６に終了報告を行うことと、転送の対象がメモリ４２ではなくキャッシュメモリ３６であることと、転送の方向が双方向であることだけである。

【０１１２】（ｖｉ）コントローラコマンド（コマンドリンク）の終了処理

次に、図１４を用いて、上述した（ｉｖ）ディスクコマンドの終了報告，（ｖ）ＤＭＡコマンドの終了報告，および下記するパリティ生成コマンドの終了報告を受信した時の、下位ディスクアレイ制御部３０３のディスクコマンド制御部３０３２における処理について説明する。図１４は、本発明の一実施形態によるディスクアレイ制御システムにおけるＭＰＵの下位ディスクアレイ制御部のディスクコマンド制御部の終了報告受信時の処理について説明するフローチャートである。

【０１１３】ステップ１２５１において、ディスクコマンド制御部３０３２は、コマンド終了報告を受信する。次に、ステップ１２５２において、ディスクコマンド制御部３０３２は、そのコマンドに該当するコントローラコマンドが「コマンドペンディング状態」であるが否かを判断し、そのコマンドに該当するコントローラコマンドが「コマンドペンディング状態」であるならば、ステップ１２５３に進み、そうでないならば、ステップ１２５５に進む。

【０１１４】ステップ１２５３において、ディスクコマンド制御部３０３２は、もし、そのフェーズが最終フェーズであるならば、起動に先立ち、コマンドのペンディングフラグを変更して、「コマンドペンディング状態」を解除する。即ち、「キャッシュロード＆リードコマンド」の処理において、図１１のステップ１２２６によるコマンドペンディング状態であるとすると、次の「キャッシュリードコマンド」は、第２フェーズであり、最終フェーズであるため、「コマンドペンディング状態」を解除する。しかしながら、第２フェーズに続いて第３フェーズがある場合には、「コマンドペンディング状態」の解除は行わない。

【０１１５】次に、ステップ１２５４において、ディスクコマンド制御部３０３２は、引き続き、コントローラコマンドの第２フェーズ，若しくは第３フェーズを起動する。即ち、図１１の（Ａ）からステップ１２２２に戻り、ステップ１２４９の前の（Ｂ）までの処理を行う。例えば、「キャッシュロード＆リードコマンド」の処理において、図１１のステップ１２２６によるコマンドペンディング状態であるとすると、ステップ１２２７の「キャッシュリードコマンド」を起動する。

【０１１６】コントローラコマンドが「コマンドペンディング状態」でないときには、ステップ１２５５において、ディスクコマンド制御部３０３２は、そのコントローラコマンドの終了でコマンドリンクに登録されたコントローラコマンドがすべて終了したが否かを判断する。

【０１１７】そのコントローラコマンドの終了でコマンドリンクに登録されたコントローラコマンドがすべて終了したならば、ステップ１２５６において、ディスクコマンド制御部３０３２は、コマンドＩ／Ｆ制御部３０３１に通知し、ホストコマンドキューから当該コマンドリンクを削除する。

【０１１８】次に、ステップ１２５７において、ＰＣＩ　Ｉ／Ｆ制御部３０１は、ＰＣＩＩ／Ｆコントローラ３２１内部のＤｏｏｒＢｅｌｌレジスタ３２１３に終了ステータスとコマンドリンクの識別番号をライトする（図８（へ））。この動作により、ＰＣＩ　Ｉ／Ｆコントローラ３２は、ＣＰＵ１０に割り込みを発行する（図８（ト））。

【０１１９】次に、図１５を用いて、終了報告を受信した時の、ホスト計算機１のＰＣＩＩ／Ｆドライバ１０４における処理について説明する。図１５は、本発明の一実施形態によるディスクアレイ制御システムにおけるホ

スト計算機のＰＣＩ　Ｉ／Ｆドライバにおける終了報告処理について説明するフローチャートである。

【０１２０】ステップ１１１１において、ホスト計算機１のＰＣＩ　Ｉ／Ｆドライバ１０４は、ＰＣＩ　Ｉ／Ｆコントローラ３２からＣＰＵ１０への割り込みを受信する。ステップ１１１２において、ＰＣＩ　Ｉ／Ｆドライバ１０４は、ＤｏｏｒＢｅｌｌレジスタ３２１３をリードし、ステップ１１１３において、コマンドリンクの識別番号を獲得し、ステップ１１１４において、終了ステータスを確認する。次に、ステップ１１１５において、ＰＣＩ　Ｉ／Ｆドライバ１０４は、コマンドリンク識別番号と終了ステータスを、上位ディスクアレイ制御ドライバ１０３の論理コマンドアドレス変換部１０３１にコマンドＩ／Ｆ１０３４経由で報告する。論理コマンドアドレス変換部１０３１は、コマンドリンクの終了処理を行い、ディスクアレイのリードコマンドの終了報告をファイルシステム１０２に報告する。

【０１２１】さらに、ＯＳ１０１は、アプリケーション１００にディスクアレイのリードコマンドの終了を報告し、コントローラコマンドの全処理が終了する。

【０１２２】（ｂ）ライト動作

次に、図１６を用いて、本発明の一実施形態による内蔵型ハイブリッドアレイ構成のディスクアレイ制御システムにおけるディスクアレイのライト動作について説明する。図１６は、本発明の一実施形態によるディスクアレイ制御システムにおけるディスクアレイのライト処理の説明図である。

【０１２３】ここで、アプリケーション１００からディスクアレイ装置２へのライト要求が発行されたとする。主記憶メモリ１１の領域１１１のライトデータＢをライトする。ライトデータＢは、ディスクアレイ装置２のストライプサイズより大きく、図１６のように、ディスクアレイ２１ａ，２１ｂ，２１ｃの異なる３つのストライプｓａ，ｓｂ，ｓｃにまたがるデータブロックａ，ｂ，ｃにより構成されるとする。

【０１２４】本実施形態では、ディスクアレイへのデータのライトは、ディスクアレイコントローラ３のキャッシュメモリ３６にデータをライトすることで処理を終了するものとする。また、キャッシュメモリ３６からディスクへの書き戻しは、非同期的に実施し、これを「デステージ動作」と称し、後述する。

【０１２５】ライト動作の処理フローは、上述した「（ａ）リード動作」とほとんど同一であるので、相違点についてのみ説明する。

【０１２６】（ｉ）　コマンドリンクの発行

アプリケーション１００がライトデータＢのライト要求を発行する。ファイルシステム１０２は、このライト要求を受信し、ディスクアレイ２１への論理ライトコマンドを上位ディスクアレイ制御ドライバ１０３に発行する。

【０１２７】次に、図７に示す処理フローに従って、上位ディスクアレイ制御ドライバ１０３は、コマンドリンクを発行する。

【０１２８】図７のステップ１１００において、上位ディスクアレイ制御ドライバ１０３の論理コマンドアドレス変換部１０３１は、論理ライトコマンドを受信する。

【０１２９】ステップ１１０１において、論理コマンドアドレス変換部１０３１は、受信した論理ライトコマンドの論理アドレスを、ディスクアレイを構成する少なくとも１台以上の個々のディスクのアドレスに変換する。上位ディスクアレイ制御ドライバ１０３の論理コマンドアドレス変換部１０３１は、図１６に示すように、ライトデータＢがディスク２１ａ，２１ｂ，２１ｃのストライプａ，ｂ，ｃに格納するデータブロックａ，ｂ，ｃにより構成されていることを認識する。

【０１３０】次に、ステップ１１０２において、キャッシュ管理部１０３２は、データブロックａ，ｂ，ｃ毎に、ディスクアレイコントローラ３のキャッシュメモリ３６のヒットミス判定を実施する。図１６に示す例においては、データｂはヒット、データａ，ｃはミスヒットであったとする。

【０１３１】ステップ１１０３において、ヒット，ミスヒットの判定を行い、ヒット時には、ステップ１１０４において、キャッシュの割当は不要であるため、キャッシュ管理部１０３２は、「キャッシュライトコマンド」を生成する。

【０１３２】ミスヒット時には、ステップ１１０５において、キャッシュ管理部１０３２は、キャッシュメモリ３６にデータをライトする領域を割り当てる。そして、ステップ１１０６において、キャッシュ管理部１０３２は、ストライプの一部にデータをライトするデータブロックｃのような場合には、一旦ディスク装置２１から当該ストライプを全部キャッシュメモリ３６にロードした後、一部データをキャッシュメモリ３６上で更新する処理を行う。この処理を行うため、キャッシュ管理部１０３２は、「キャッシュロード＆ライトコマンド」を生成する（１１０６）。データブロックａのようにストライプ全体をライトする場合は、「キャッシュライトコマンド」を生成する。

【０１３３】なお、キャッシュメモリ３６へのデータブロックのライトによりキャッシュメモリ３６とディスクのデータの不整合が発生する。これを識別するため、当該ストライプのキャッシュには「Ｄｉｒｔｙ」の識別子を設定する。なお、後のデステージ処理において、キャッシュメモリ３６とディスク２１のデータが一致した際には、この識別子を「Ｃｌｅａｎ」に設定する。

【０１３４】ステップ１１０７において、キャッシュ管理部１０３２は、生成したコントローラコマンドについて、同一論理コマンド毎にコマンドリンクを生成し、主記憶メモリの特定の領域に格納する。

【０１３５】ステップ１１０８において、キャッシュ管理部１０３２は、ステップ１１０２～１１０７の処理を、論理コマンドの扱うライトデータを構成する全てのデータブロックについて実施する。

【０１３６】図１６に示す例では、データブロックａに対応するキャッシュライトコマンドと、データブロックｂに対応するキャッシュライトコマンドと、データブロックｃに対応するキャッシュロード＆ライトコマンドと、の３つのコマンドがリンクされ、キャッシュメモリ３６に格納される。

【０１３７】この例の場合、データブロックｃは、ストライプｃの一部データのみをライトするので、キャッシュロード＆ライトコマンドには、ディスク２１ｃ上のストライプｃの先頭アドレスと、データブロックｃの先頭アドレスの両者を格納している。

【０１３８】次に、ステップ１１０９において、コマンドＩ／Ｆ制御部１０３は、ＰＣＩＩ／Ｆドライバ１０４に指示し、コマンドリンクをディスクアレイコントローラ３に発行する。

【０１３９】以下、「コマンドリンク発行パケット」のディスクアレイコントローラへの送信は、上述した「（ａ）リード動作」の「（ｉ）コマンドリンク発行パケット」と同様である。

【０１４０】（ｉｉ）コマンドリンクの受信

「（ａ）リード動作」の「（ｉｉ）コマンドリンクの受信」と同様である。

【０１４１】（ｉｉｉ）コントローラコマンド（コマンドリンク）の起動処理

次に、図１１及び図１２を用いて、コマンドＩ／Ｆ制御部３０３１のホストコマンドキューにコマンドリンクがキューイングされた時の、下位ディスクアレイ制御部３０３における処理について説明する。

【０１４２】「（ａ）リード動作」と同様に、コマンド管理部３０３６は、ステップ１２２０において、ホストコマンドキューからコマンドリンクを一つ取り出し、ステップ１２２１において、そのコマンドリンクに登録されたコントローラコマンドを一つ取り出し、ステップ１２２１において、コントローラコマンドを解析する。

【０１４３】図１６の例では、コントローラコマンドとして、データブロックｃに対する「キャッシュロード＆ライトコマンド」と、データブロックａ，ｂに対する「キャッシュライトコマンド」が使用される。データブロックａ，ｂ，ｃの順にコントローラコマンドの起動処理を行う。以下、この２つのコマンドについてコマンド管理部の動作を図１１及び図１２を用いて説明する。

【０１４４】図１１のステップ１２３０において、コマンド管理部３０３６が、「キャッシュロード＆ライトコマンド」と判断すると、ステップ１２３１に進む。

【０１４５】ステップ１２３１において、コマンド管理部３０３６は、第１フェーズの処理として、当該ストライプの全データを上位ディスクアレイ制御ドライバの指定したキャッシュメモリ３６のアドレスにディスク２１からデータをロードするためのディスクロードコマンドを生成するなお、本実施形態では、必ずキャッシュへのロードは、ストライプ単位であるとする。よって、データブロックｃは、アプリケーションからのリード要求はストライプの一部ではあるが、上位ディスクアレイ制御ドライバの指示に基づき、ディスク２１ｃからキャッシュメモリ３６へのロードは、ストライプｓｃ全部をロードすることになる。

【０１４６】ステップ１２３２において、コマンド管理部３０３６は生成したディスクコマンドをディスクコマンド制御部３０３２のディスクコマンドキューにキューイングする。

【０１４７】さらに、ステップ１２３３において、コマンド管理部３０３６は、このコントローラコマンドを「コマンドペンディング状態」に設定し、このコントローラコマンドの処理は一旦中断し、コントローラコマンドの終了を待つものである。次に、コマンド管理部３０３６は、図１２のステップ１２４９に進み、次のコントローラコマンドの処理を続行する。ディスクロードコマンド終了後は、同図の点線に従い、以下に示すキャッシュライトコマンドの処理を実施する。

【０１４８】図１１のステップ１２３４において、コマンド管理部３０３６が、「キャッシュライトコマンド」と判断すると、ステップ１２３５に進む。

【０１４９】ステップ１２３５において、コマンド管理部３０３６は、ディスクからデータをキャッシュメモリ３６にロードする必要なく、主記憶メモリ１１のデータキャッシュメモリ３６にＤＭＡ転送するためのキャッシュライトＤＭＡコマンドを生成する。

【０１５０】そして、ステップ１２３６において、コマンド管理部３０３６は、キャッシュライトＤＭＡコマンドをＤＭＡコマンドキューにキューイングし、キャッシュライトコマンドの終了を待つものである。

【０１５１】（ｉｖ）ディスクコマンドの処理

「（ａ）リード動作」の「（ｉｖ）ディスクコマンドの処理」と同様である。

【０１５２】（ｖ）ＤＭＡコマンドの処理

「（ａ）リード動作」の「（ｖ）ＤＭＡコマンドの処理」と同様である。

【０１５３】（ｖｉ）コントローラコマンド（コマンドリンク）の終了処理

「（ａ）リード動作」の「（ｖｉ）コントローラコマンド（コマンドリンク）の終了処理」と同様である。「コマンドペンディング状態」のコントローラコマンドは、第２フェーズの処理を実施し、コマンドリンクに登録された全てのコントローラコマンドの処理を実行し終わり、アプリケーションに「（ａ）リード動作」と同様に終了報告がなされれば、この論理ライトコマンドの処理

はすべて終了する。

【０１５４】（ｃ）デステージ動作

本実施形態においては、上述したように、データのライト処理は、キャッシュメモリ３６に対して実行し、ディスクへの書き戻しは、この処理から遅延させて非同期に実行する。この書き戻し処理を、「デステージ処理」と称する。

【０１５５】デステージ処理の起動は、ホスト計算機１の上位ディスクアレイ制御ドライバ１０３が行う。キャッシュ管理部１０３２は、キャッシュの利用状況を考慮し、例えば、ある一定以上のキャッシュに「Ｄｉｒｔｙ」なブロックが格納されたときや、ライトが行われ、ある一定時間経過したとき等に、キャッシュの「Ｄｉｒｔｙ」ブロックのデステージ処理を起動する。

【０１５６】デステージ処理は、処理効率をあげるため、まとめ処理を行う等、様々なアルゴリズムを実装できるが、本実施形態では、２つの単純な例で説明する。それ以上の複雑な実装も、その応用として同様に実現できるものである。

【０１５７】ここで、図１７及び図１８を用いて、デステージ動作について説明する。

【０１５８】（ｉ）コマンドリンクの発行

ここで、図１７を用いて、１つのＤｉｒｔｙブロックＣ（新データストライプ）のデステージ処理の動作について説明する。図１７は、本発明の一実施形態によるディスクアレイ制御システムにおけるディスクアレイの第１のデステージ処理の説明図である。

【０１５９】本例においては、１つのＤｉｒｔｙブロックＣ（新データストライプ）３６７をデステージする場合について説明する。

【０１６０】上位ディスクアレイ制御ドライバ１０３のキャッシュ管理部１０３２は、ＤｉｒｔｙブロックＣのデステージを行うことを決定すると、上記同様、コントローラコマンドを生成する。

【０１６１】この例では、新パリティストライプ３６７を生成するために必要な旧データストライプ３６８と旧パリティストライプ３７０がキャッシュメモリ３６上に存在しないとすると、（１）これらのストライプ３６８，３７０をリードし、（２）新データストライプであるＤｉｒｔｙブロックＣ３６７の３者の排他的論理和（以下、「ＸＯＲ」と略記する）を演算することで新パリティストライプ３６９を生成し、（３）ＤｉｒｔｙブロックＣ３６７および新パリティストライプ３６９の両者をディスク２１にライトする、という３フェーズを実行する必要がある。

【０１６２】この処理のためのコントローラコマンドとして、「キャッシュロード＆デステージコマンド」を生成する。「キャッシュロード＆デステージコマンド」については、図１２を用いて後述する。「キャッシュロード＆デステージコマンド」には、上記（１）から（３）の処理に必要な、キャッシュメモリのアドレス情報，ディスクのアドレス情報，コントローラコマンドの識別番号等の必要な情報をすべて格納する。

【０１６３】また、図１８を用いて、ディスクアレイの同一パリティグループを構成する全てのデータブロックをまとめてデステージするデステージ処理の動作について説明する。図１８は、本発明の一実施形態によるディスクアレイ制御システムにおけるディスクアレイの第２のデステージ処理の説明図である。

【０１６４】本例においては、ディスクアレイのディスク数は５台で、パリティグループを４つのデータと１つのパリティで構成する４Ｄ＋１Ｐ構成としている。

【０１６５】上位ディスクアレイ制御ドライバ１０３のキャッシュ管理部１０３２は、Ｄｉｒｔｙブロック（新データストライプ）Ｄ０，Ｄ１，Ｄ２，Ｄ３（３７１，３７２，３７３，３７４）のデステージを行うことを決定すると、新パリティストライプＰは、

Ｐ＝Ｄ０＋Ｄ１＋Ｄ２＋Ｄ３　（ただし、＋は、ＸＯＲ演算を示す。）

として生成できるので、（１）パリティストライプＰの生成、（２）全新データおよび新パリティストライプのデステージ、の２フェーズを実行する必要がある。

【０１６６】この処理のためのコントローラコマンドとして、「キャッシュデステージコマンド」を生成する。「キャッシュデステージコマンド」については、図１２を用いて後述する。「キャッシュデステージコマンド」には、上記（１）から（２）の処理に必要な、キャッシュメモリのアドレス情報，ディスクのアドレス情報，コントローラコマンドの識別番号等の必要な情報をすべて格納する。

【０１６７】以上の処理をすべて、もしくは一部の対象Ｄｉｒｔｙブロックに対し実施し、コマンドリンクを生成し、コマンド発行パケットを上記同様ディスクアレイコントローラ３に送信する。

【０１６８】（ｉｉ）コマンドリンクの受信

「（ａ）リード動作」の「（ｉｉ）コマンドリンクの受信」と同様である。

【０１６９】（ｉｉｉ）コントローラコマンド（コマンドリンク）の起動処理

図１１に示すように、「（ａ）リード動作」と同様に、コマンド管理部３０３６は、ステップ１２２０において、ホストコマンドキューからコマンドリンクを一つ取り出し、ステップ１２２１において、そのコマンドリンクに登録されたコントローラコマンドを一つ取り出し、ステップ１２２２において、コントローラコマンドを解析する。

【０１７０】図１７の例では、コントローラコマンドとして、上述したように、「キャッシュロード＆デステージコマンド」が使用される。また、図１８の例では、コントローラコマンドとして、上述したように、「パリテ

ィ生成＆デステージコマンド」が使用される。

【０１７１】次に、図１２を用いて、これらのコントローラコマンドの処理におけるコマンド管理部３０３６の動作について説明する。

【０１７２】図１２のステップ１２３６において、コマンド管理部３０３６が、「キャッシュロード＆デステージコマンド」と判断すると、ステップ１２３７に進む。ステップ１２３７において、コマンド管理部３０３６は、第１フェーズの処理として、旧データストライプ３６８と、旧パリティストライプ３７０の両者をディスク２１ｃからキャッシュメモリ３６にロードするためのディスクリードコマンドを生成する。ステップ１２３８において、コマンド管理部３０３６は、生成したディスクコマンドを、ディスクコマンド制御部３０３２のディスクコマンドキューにキューイングする。

【０１７３】さらに、ステップ１２３９において、コマンド管理部３０３６は、このコントローラコマンドを「コマンドペンディング状態」に設定し、このコントローラコマンドの処理は一旦中断し、コントローラコマンドの終了を待つものである。ディスクコマンド終了後は、第２フェーズとして、同図の点線に従い、以下に示す「パリティ生成＆デステージコマンド」の処理を実施する。

【０１７４】図１２のステップ１２４２において、コマンド管理部３０３６が、「パリティ生成＆デステージコマンド」と判断すると、ステップ１２４３に進む。

【０１７５】ステップ１２４３において、コマンド管理部３０３６は、第１フェーズの処理として（キャッシュロード＆デステージコマンドから引き継いだときは、第２フェーズの処理として）、新パリティストライプの生成のために必要なデータストライプのＸＯＲ演算を実行するための「パリティ生成コマンド」を生成する。図１７に示した第１の例では、新データストライプ３６７と、旧データストライプ３６８と、旧パリティ３７０とをＸＯＲ演算して、新パリティ３６９を生成するための「パリティ生成コマンド」を生成する。また、図１８に示した例では、新データストライプＤ０，Ｄ１，Ｄ２，Ｄ３（３７１，３７２，３７３，３７４）のＸＯＲ演算を実行して、新パリティ３７５を生成するための「パリティ生成コマンド」を生成する。

【０１７６】ステップ１２４４において、コマンド管理部３０３６は、生成したパリティ生成コマンドをパリティ生成コマンド制御部３０３３のパリティ生成コマンドキュー３０３３１にキューイングする。

【０１７７】さらに、ステップ１２４５において、コマンド管理部３０３６は、このコントローラコマンドを「コマンドペンディング状態」に設定し、このコントローラコマンドの処理は一旦中断し、コントローラコマンドの終了を待つものである。ディスクコマンド終了後は、第２フェーズとして、同図の点線に従い、以下に示す「キャッシュデステージコマンド」の処理を実施する。

【０１７８】「キャッシュデステージコマンド」は、ホスト計算機１の上位ディスクアレイ制御ドライバ１０３からは発行されないコントローラコマンドである。キャッシュデステージコマンドは、キャッシュロード＆デステージコマンド、及びパリティ生成＆デステージコマンドから引き継いで、第２，若しくは第３フェーズの処理として実行され、コントローラコマンドに登録されたＤｉｒｔｙな新データストライプもしくは生成した新パリティストライプのディスクへのライトを行うものである。

【０１７９】図１２のステップ１２４６において、コマンド管理部３０３６が、「キャッシュデステージコマンド」と判断すると、ステップ１２４７に進む。

【０１８０】ステップ１２４７において、コマンド管理部３０３６は、ディスクライトコマンドを生成する。図１７の例では、新データストライプＣ３６７および新パリティストライプ３６９をデステージするため、これら全てのディスクライトコマンドを生成する。また、図１８に示す例では、新データストライプＤ０，Ｄ１，Ｄ２，Ｄ３（３７１，３７２，３７３，３７４）および新パリティストライプＰ３７５をデステージするため、これら全てのディスクライトコマンドを生成する。

【０１８１】ステップ１２４８において、コマンド管理部３０３６は、これらのディスクライトコマンドをディスクコマンドキューにキューイングし、全ディスクライトコマンドの終了を待つものである。

【０１８２】（ｉｖ）ディスクコマンドの処理

「（ａ）リード動作」の「（ｉｖ）ディスクコマンド」と同様である。

【０１８３】（ｖ）　パリティ生成コマンドの処理

次に、図１９を用いて、パリティ生成コマンドキュー３０３３１にコマンドがキューイングされた時の、下位ディスクアレイ制御部３０３のパリティ生成コマンド制御部３０３３における処理について説明する。図１９は、本発明の一実施形態によるディスクアレイ制御システムにおけるＭＰＵの下位ディスクアレイ制御部のパリティ生成コマンド制御部の処理について説明するフローチャートである。

【０１８４】ステップ１２７０において、パリティ生成コマンド制御部３０３３は、パリティ生成コマンドキュー３０３３１から一つのコマンドを取り出す。ステップ１２７１において、パリティ生成コマンド制御部３０３３は、パリティ生成コマンドを起動する。パリティ生成コマンド制御部３０３３は、処理を停止し、スリープする。この間、ＭＰＵ３０は他の処理を実行できる。パリティ生成回路３５は、キャッシュメモリ３６からＸＯＲ演算対象データをリードし、演算結果を新パリティとしてキャッシュメモリ３６にライトする。

【０１８５】パリティ演算が終了すると、パリティ生成回路３５は終了割り込みをＭＰＵ３０に発行する。ステップ１２７２において、ＭＰＵ３０は、終了割り込みを受信する。そして、パリティ生成コマンド制御部３０３３が処理を再開する。ステップ１２７３において、パリティ生成コマンド制御部３０３３は、終了報告をコマンド管理部３０３６に対し行う。ステップ１２７４において、パリティ生成コマンド制御部３０３３は、この終了したパリティ生成コマンドを、パリティ生成コマンドキュー３０３３１から削除する。

【０１８６】ステップ１２７４において、パリティ生成コマンド制御部３０３３は、パリティ生成コマンドキュー３０３３１に未起動のコマンドがあるか否かを判断し、パリティ生成コマンドキュー３０３３１に未起動のコマンドがある場合には、上記の起動処理を実行し、なければ処理を終了する。

【０１８７】（ｖｉ）コントローラコマンド（コマンドリンク）の終了処理

「（ａ）リード動作」の「（ｖｉ）コントローラコマンド（コマンドリンク）の終了処理」と同様である。

【０１８８】「コマンドペンディング状態」のコントローラコマンドは、第２，第３フェーズの処理を実施し、コマンドリンクに登録された全てのコントローラコマンドの処理を実行し終わったなら、上記「（ａ）リード動作」と同様の方法で、上位ディスクアレイ制御ドライバ１０３のキャッシュ管理部１０３２に終了報告を行い、キャッシュ管理部１０３２は、キャッシュメモリ３６とディスク２１のデータが一致したので「Ｄｉｒｔｙ」識別子を「Ｃｌｅａｎ」に設定する。以上でデステージ処理は終了する。

【０１８９】（ｄ）縮退動作

ディスクアレイを構成するディスク２１の内の１つのディスクが故障して、データストライプが失われた場合、パリティストライプから故障ディスクのデータを再現することができる。この動作を、「縮退動作」と称する。

【０１９０】図１８に示したような４Ｄ＋１Ｐ構成のディスクアレイの場合、ディスク２１ａ，２１ｂ，２１ｃ，２１ｄ，２１ｅのうち、例えば、ディスク２１ａが故障したとすると、各々に格納されるデータストライプＤ０，Ｄ１，Ｄ２，Ｄ３のうち、データストライプＤ０が失われる。しかしながら、データストライプＤ０は、パリティストライプＰを用い、以下の式に基づき再現できる。

【０１９１】Ｄ０＝Ｄ１＋Ｄ２＋Ｄ３＋Ｐ　（但し、＋はＸＯＲ演算を表す。）

また、ディスクアレイを構成するディスク２１の内の１つのディスクが故障した状態で、新規なデータストライプＤ０ｎｅｗをライトするときは、新パリテイＰ１を生成する必要がある。そこで、以下の式に基づき、ディスク２１ｂ，２１ｃ，２１ｄからデータストライプＤ１，Ｄ２，Ｄ３をリードし、新規なデータストライプＤ０ｎｅｗとのＸＯＲ演算の実行により、新パリテイＰ１を生成し、新パリテイＰ１のみをディスク２１ｅにライトする。

【０１９２】Ｐ１＝Ｄ０ｎｅｗ＋Ｄ１＋Ｄ２＋Ｄ３

（但し、＋はＸＯＲ演算を表す。）

これらの縮退動作については、図１１及び図１２に示す処理フロー中には記載していないが、すべて上述した動作の組み合わせで実現できる。または、新規にコントローラコマンドを用意することもできるが、やはり上記動作の応用として容易に実現できるものである。

【０１９３】以上の説明したように、ディスクアレイコントローラ３をホスト計算機１に内蔵し、ホストＩ／ＦとしてＰＣＩ　Ｉ／Ｆ３２を用いて、上位ディスクアレイ制御ドライバ１０３と下位ディスクアレイ制御部３０３との間のコマンド転送や、主記憶メモリ１１とキャッシュメモリ３６間のデータ転送を行うことができる。

【０１９４】このとき、第２ホストＩ／ＦであるＳＣＳＩは用いていないので、ＳＣＳＩコントローラ３３１は、ディスク用のＳＣＳＩコントローラとして使用することができる。

【０１９５】以上説明したように、ディスクアレイ処理を、ホスト計算機１のＣＰＵ１０とディスクコントローラ３のＭＰＵ３０とで分散した「ハイブリッドアレイ」構成とすることにより、処理負荷の重いディスクアレイの処理に、処理能力の高いＣＰＵ１０で処理負荷の重い上位ディスクアレイ制御を実行し、処理能力の低いＭＰＵ３０で処理負荷の軽い下位ディスクアレイ制御を実施することができ、従来のディスクアレイコントローラのＭＰＵが性能ボトルネックになることによる限界性能を越える高性能化を実現することができるものである。

【０１９６】本実施形態によれば、ディスクアレイコントローラに低価格なＭＰＵを使用したディスクアレイ制御システムにおいて、ＭＰＵの能力で性能が制限されることなく高性能なディスクアレイ制御が実現できる。

【０１９７】次に、図２０～図２２を用いて、本発明の第２の実施形態によるディスクアレイ制御システムについて説明する。

【０１９８】［構成の説明］最初に、図２０を用いて、本発明の第２の実施形態による外付け型ハイブリッドアレイ構成のディスクアレイ制御システムの全体構成について説明する。図２０は、本発明の第２の実施形態によるディスクアレイ制御システムのハードウエアの全体構成のブロック図である。なお、図１と同一符号は、同一部分を示しており、以下の説明においては、図１との相違点を中心に説明する。

【０１９９】ホスト計算機１は、図１に示した構成に加え、ＳＣＳＩ１５を制御するＳＣＳＩコントローラ１５１と、ＳＣＳＩ１５を接続するＳＣＳＩコネクタ１５２を備えている。

【０２００】ディスクアレイ装置２は、ホスト計算機１とは独立した別の筐体を備える。この構成を、「外付け型」と称する。ディスクアレイ装置２は、図１の構成に加え、電源２２と、クロック発生器２３と、ホストＩ／ＦとしてＳＣＳＩを選択するＳＣＳＩ選択手段２５とを備えている。また、ディスクアレイ装置２は、筐体内の温度を監視する温度監視手段２８１と、ディスクの挿抜を検出する挿抜検出手段２８２と、ＬＥＤの点灯制御を行うＬＥＤ制御手段２８３と、電源２２やＦＡＮ２００の故障を検出する故障検出手段２８４と、温度監視手段２８１と挿抜検出手段２８２とＬＥＤ制御手段２８３と故障検出手段２８４を制御し、筐体に発生する全ての異常検出を行う筐体異常検出手段２８とを備えている。なお、ディスクアレイ装置２は、さらに、ハードアレイ選択手段２９を備えているが、これについては、第３の実施形態において説明する。

【０２０１】ディスクアレイコントローラ３は、図１の構成に加え、使用するホストＩ／Ｆを選択するホストＩ／Ｆモード選択手段３７ａと、筐体の異常を監視する筐体異常監視手段４３とを備えている。なお、ディスクアレイコントローラ３は、ディスクアレイモード選択手段４４を備えているが、これについては、第３の実施形態において説明する。

【０２０２】次に、図２１を用いて、本発明の第２の実施形態による外付け型ハイブリッドアレイ構成のディスクアレイ制御システムの立体的構成について説明する。図２１は、本発明の第２の実施形態によるディスクアレイ制御システムのハードウエアのブロック斜視図である。なお、図２０と同一符号は、同一部分を示している。

【０２０３】ディスクアレイ装置２は、さらに、ディスクアレイコントローラ３を搭載するためのコネクタ２４と、バックボード２９０と、バックボード２９０とディスクアレイコントローラ３を接続するバックボードＩ／Ｆコネクタ２５１と、筐体内部の空調を行うＦＡＮ２００と、ディスクや電源やＦＡＮの状態を表示するＬＥＤ２１３ａ，ｂ，ｃ…，２２３，２０１を備えている。

【０２０４】ディスクアレイ装置２のバックボード２９０は、カードエッジ３１を接続するコネクタ２４を備えており、ディスクアレイコントローラ２を接続する。ブラケット２７ａは、ディスクアレイ装置２の筐体に固定されている。ブラケット２７ａは、ホスト計算機１とＳＣＳＩケーブル１７に接続されている。筐体には、ブラケット２７ａの逆側の横辺を固定するための、支え金具２７ｂが設けられている。

【０２０５】ディスクアレイコントローラ３は、さらに、ディスクアレイ装置２のバックボード２９０と接続するバックボードＩ／Ｆコネクタ３７３とＰＣＩ　Ｉ／Ｆ３２に接続するためのカードエッジ３１を備えている。

【０２０６】［動作の説明］

（１）ホストＩ／Ｆの選択

図２２を用いて、ホストＩ／Ｆモード選択手段３７ａによるホストＩ／Ｆの選択について説明する。図２２は、本発明の第２の実施形態によるディスクアレイ制御システムのホストＩ／Ｆモード選択手段の構成を示す回路図である。

【０２０７】ディスクアレイコントローラ３のホストＩ／Ｆモード選択手段３７ａは、どちらのホストＩ／Ｆを選択しているかを示すＩＦ＿Ｍｏｄｅ信号３７２をＭＰＵ３０に送出する。ＩＦ＿Ｍｏｄｅ信号３７２は、抵抗で電源にプルアップされている。ホストＩ／Ｆモード選択手段３７ａは、バックボードＩ／Ｆコネクタ３７３を備えている。

【０２０８】ディスクアレイ装置内部のバックボード２９０は、ＳＣＳＩ選択手段２５を備えている。ＳＣＳＩ選択手段２５のバックボードＩ／Ｆコネクタ２５１は、バックボードＩ／Ｆコネクタ３７３と対になるものであり、グランド線に接地してされている。

【０２０９】図２１に示したように、ディスクアレイコントローラ３をディスクアレイ装置２のバックボード２９０に実装すると、ホストＩ／Ｆモード選択手段３７ａのバックボードＩ／Ｆコネクタ３７３と、ＳＣＳＩ選択手段２５のバックボードＩ／Ｆコネクタ２５１は接続する。この結果、ＩＦ＿Ｍｏｄｅ信号３７２は、信号レベルが”Ｌ”になる。一方、もし、両者が接続していないときは、バックボードＩ／Ｆコネクタ３７３がオープンになり、プルアップ抵抗により、ＩＦ＿Ｍｏｄｅ信号３７２は、信号レベルが”Ｈ”になる。従って、バックボード２９０との接続しているかどうかにより、図４に示したように、ＩＦ＿Ｍｏｄｅ信号３７２が変化することになる。

【０２１０】以上のように、ディスクアレイコントローラ３を、ホスト計算機１から独立したディスクアレイ装置２に内蔵している「外付け型」のときは、ホストＩ／Ｆとして、ＳＣＳＩ３３を選択し、そうでないとき、すなわち、ホスト計算機１に内蔵している「内蔵型」のときは、ホストＩ／Ｆとして、ＰＣＩ　Ｉ／Ｆ３２を選択することができる。

【０２１１】本実施形態においては、ディスクアレイコントローラ３がディスクアレイ装置２に内蔵され、バックボード２９０に接続しているものとする。

【０２１２】ディスクアレイ装置２の電源が投入されると、ＩＦ＿Ｍｏｄｅ信号３７２が信号レベル”Ｌ”でＭＰＵ３０に入力し、ＭＰＵ３０で動作する図２に示したホストＩ／Ｆモード制御部３０５はこれを認識し、図５のステップ１００９において、ホストＩ／Ｆとして、ＳＣＳＩ３３を選択し、初期化部３０３５は、ディスクアレイコントローラ３の初期化を実施する。

【０２１３】（２）　ディスクアレイコントローラの実

装
次に、図21を用いて、ディスクコントローラ3のディスクアレイ装置2への実装の方法について説明する。
【0214】ディスクアレイコントローラ3は、PCI I／F32に接続するためのカードエッジ31を備えている。ディスクアレイ装置2のバックボード290は、カードエッジ31を接続するコネクタ24を備えており、ディスクアレイコントローラ2を接続する。ホストI／Fとして使用するSCSI33は、ディスクアレイコントローラ2の横辺に備えるブラケット27aに固定したSCSIコネクタ332を用い、ホスト計算機1とSCSIケーブル17によって接続する。このブラケット27aは、ディスクアレイ装置2の筐体に固定されている。また、ディスクアレイ装置2の筐体には、ブラケット27aの逆側の横辺を固定するための支え金具27bを設けてある。カードエッジ31用コネクタ24，ブラケット27a，支え金具27bの3点によって、ディスクアレイコントローラ3は、ディスクアレイ装置2の筐体に固定されている。
【0215】また、ディスクアレイコントローラ3への電力や動作用クロックの供給は、バックボード290からPCI I／F32のカードエッジ31を介して行われる。
【0216】バックボード290上には、クロック発生器23を備えている。また、ディスクアレイ装置2には、電源22を備えている。クロック発生器23が送出するクロック信号231と、電源22が送出する電力を伝達する電源線222は、バックボード290に備えてあり、コネクタ24からPCI I／F32のカードエッジ仕様にあわせ接続する。
【0217】以上のような構成により、ディスクアレイコントローラ3をホスト計算機1に内蔵してPCI I／F32をホストI／Fとして使用するときと全く同じ方法で、ディスクアレイコントローラ3に電力とクロック信号を供給できる。
【0218】（3）筐体異常の検出と制御
ディスクアレイ装置2は、ホスト計算機1とは独立しているので、上述したように、ディスクアレイ装置2は、専用の電源22を備えている。また、装置内部の空調用のファン200を備えている。また、図21には記載していないが、電源もファンも複数台備えることで耐故障性を持つことができる。
【0219】これらの各部位の故障の有無を、ディスクアレイコントローラは監視する必要がある。そこで、ディスクアレイ装置2は、故障検出手段284を備えており、故障検出手段284が、電源22，ファン200の異常を検出する。
【0220】また、電源22，ファン200は、冗長構成の場合、オンライン中に交換することができる。そこで、どの電源22，ファン200が故障しているかをディスクアレイ装置2のユーザに視覚的に報告する必要があるので、電源22及びファン200は、それぞれ、LED223，201を備えている。LED制御手段283は、LED223，201の点灯制御を行うものである。
【0221】また、ディスクアレイ装置2は、ディスクの故障に対しても耐故障性を備えている。ディスクが故障した際には、縮耐運転を行うことで、ディスクアレイの処理を継続可能である。ただし、同一パリティグループを構築するディスクが2台以上同時に故障すると、データを喪失してしまうので、1台のディスクが故障したら速やかにディスクを交換する必要がある。そこで、ディスクアレイ装置2のユーザには、ディスク装置21a，21b，21c，21dの状態と故障ディスクを視覚的に報告する必要があるので、ディスク装置21a，21b，21c，21dは、LED213a，…，213eを備えている。上記LED制御手段283は、LED213a，…，213eの点灯制御を行う。
【0222】また、故障したディスクが抜き取られたり、新しいディスクが挿入されたりした際には、この状態変化をディスクアレイコントローラ3は検出する必要がある。挿抜検出手段282は、ディスクの挿抜を検出する。
【0223】また、ディスクアレイ装置2の内部は何らかの異常により、温度が上昇し、装置全体に悪影響を及ぼす恐れがある。そこで、温度監視手段281は、ディスクアレイ装置内部の温度を監視する。
【0224】故障検出手段284，LED制御手段283，挿抜検出手段282，温度監視手段281は、ディスクアレイ装置2に備えた筐体異常検出手段28に接続される。筐体異常検出手段28は、ディスクアレイコントローラ2が備える筐体監視手段43に接続される。筐体異常検出手段28は、ディスクアレイ装置2の内部で発生した上記異常を検出し、筐体監視手段43に報告する。また、筐体監視手段43は、必要なLEDの点灯指示を筐体異常検出手段28経由でLED制御手段283に発行する。
【0225】ディスクアレイコントローラ3上の筐体監視手段43と、ディスクアレイ装置2の筐体異常検出手段28の接続には、上記バックボードI／Fコネクタ373，251を使用する。これにより、特別なケーブル無しにディスクアレイコントローラ3が、筐体の異常情報を検出することができる。
【0226】なお、この筐体異常の検出の方法は、第1実施形態記載のホストI／FとしてPCI I／F32を選択した際にも使用することができるものである。
【0227】（4）ディスクアレイ動作
次に、ディスクアレイの動作について説明する。本実施形態においては、ホストI／FとしてPCI I／F32ではなくSCSI33を用いるようにしているので、

ホスト計算機１の上位ディスクアレイ制御ドライバ１０３で生成したコントローラコマンドをディスクアレイコントローラ３の下位ディスクアレイ制御部３０３に送信する方法が異なるものである。そこで、コントローラコマンドの送信方法について、以下に説明する。

【０２２８】コントローラコマンドの種類と生成の方法は、上述した第１の実施形態と同様である。相違する点は、ＳＣＳＩ３３を用いる場合には、ＳＣＳＩのコマンドブロックであるＣＤＢ（Ｃｏｍｍａｎｄ　Ｄｉｓｃｒｉｐｔｅｒ　Ｂｌｏｃｋ）として実現することである。また、コマンドリンクは、ＳＣＳＩの仕様であるリンクコマンド機能を用いて実現できる。この機能は、ＣＤＢのＬｉｎｋビットを”１”にセットすることで、連続してＣＤＢを送信し、これらのリンクされたコマンドが終了した時点で終了報告をまとめて受信する。この機能を用いることで、第１の実施形態と等価なコマンドリンクを実現できる。

【０２２９】次に、コマンドリンクの発行の手順を説明する。図２に示した上位ディスクアレイ制御ドライバ１０３のキャッシュ管理部１０３２は、コントローラコマンドを生成し、主記憶メモリ１１にコマンドリンクを生成する。コマンドＩ／Ｆ制御部１０３４は、ＳＣＳＩドライバ１０５にコマンド発行を指示する。ＳＣＳＩドライバ１０５は、ＳＣＳＩコントローラ１５のコマンド起動ビットをセットする。ＳＣＳＩコントローラ１５は、インテリジェント型である場合、コマンド起動ビットのセットにより、コマンドリンクをＳＣＳＩコントローラ１５の内部のＤＭＡコントローラでとりこみ、ＳＣＳＩバスを駆動して、ディスクアレイコントローラ３に送信する。

【０２３０】ディスクアレイコントローラ３のＳＣＳＩコントローラ３３１は、このコマンドリンクを受信し、その内部のＤＭＡコントローラでメモリ４２に転送し、ＭＰＵ３０に割り込みを発行する。ＭＰＵ３０は、この割り込みを受信し、ＳＣＳＩ制御部３０２を起動する。ＳＣＳＩ制御部３０２は、受信したコマンドリンクを取り込み、下位ディスクアレイ制御部３０３のホストコマンドキューにキューイングする。以下の処理は、上述した第１の実施形態と同様である。

【０２３１】また、ホスト計算機１の主記憶メモリ１１とディスクアレイコントローラ３のキャッシュメモリ３６との間のデータ転送は、上述した第１の実施形態では、図８に示したＰＣＩ　Ｉ／Ｆコントローラ３２１の有するＤＭＡコントローラ３２１１が直接両者間を転送したが、本実施形態ではＳＣＳＩを用いるため、このような転送ができない。

【０２３２】そこで、主記憶メモリ１１とのキャッシュメモリ３６との間のデータ転送の方法について、ディスクアレイのキャッシュリードを例にとって、以下に、説明する。ここでは、キャッシュメモリ３６から主記憶メモリ１１へデータが転送される。キャッシュライト時は、方向が逆になるだけで同様である。

【０２３３】図２において、ディスク２１からデータがキャッシュメモリ３６に転送されると、コマンド管理部３０３６は、ＳＣＳＩ制御部３０２に通知する。ＳＣＳＩバスは、コマンドを受信した後、キャッシュメモリ３６にデータが準備されるまでの間ディスコネクトしており、この間、ホストＩ／ＦであるＳＣＳＩ３３は他のコマンドの受信や、他のコマンドのデータ転送に用いることができる。ＳＣＳＩ制御部３０２は、この通知を受けＳＣＳＩバスをリコネクトし、ＳＣＳＩコントローラ３３１の内部のＤＭＡコントローラを起動し、キャッシュメモリ３６からデータをＳＣＳＩコントローラ３３１内部に転送し、このデータをＳＣＳＩ３３経由でホスト計算機１のＳＣＳＩコントローラ１５に転送する。ホスト計算機１のＳＣＳＩコントローラ１５の内部のＤＭＡコントローラは、このデータを受信し、主記憶メモリ１１にＤＭＡ転送する。

【０２３４】また、コントローラコマンドの終了報告も、ＳＣＳＩ３３のステータスを用いて実現することができる。

【０２３５】なお、デステージ処理は、ホスト計算機１からコントローラコマンドが送信されるだけで、ディスクアレイ装置２内部で処理が実行されるので、ＳＣＳＩバス上でデータ転送は行われないので、上述した第１の実施形態と同様である。

【０２３６】以上の説明したように、ディスクアレイコントローラ３をホスト計算機１に内蔵し、ホストＩ／ＦとしてＰＣＩ　Ｉ／Ｆ３２を用いて、上位ディスクアレイ制御ドライバ１０３と下位ディスクアレイ制御部３０３との間のコマンド転送や、主記憶メモリ１１とキャッシュメモリ３６間のデータ転送を行うことができる。

【０２３７】このとき、第２ホストＩ／ＦであるＳＣＳＩは用いていないので、ＳＣＳＩコントローラ３３１は、ディスク用のＳＣＳＩコントローラとして使用することができる。

【０２３８】以上説明したように、ディスクアレイ処理を、ホスト計算機１のＣＰＵ１０とディスクコントローラ３のＭＰＵ３０とで分散した「ハイブリッドアレイ」構成とすることにより、処理負荷の重いディスクアレイの処理に、処理能力の高いＣＰＵ１０で処理負荷の重い上位ディスクアレイ制御を実行し、処理能力の低いＭＰＵ３０で処理負荷の軽い下位ディスクアレイ制御を実施することができ、従来のディスクアレイコントローラのＭＰＵが性能ボトルネックになることによる限界性能を越える高性能化を実現することができるものである。

【０２３９】また、第１の実施形態及び第２の実施形態において説明したように、同一のディスクアレイコントローラを用いながら、ホストＩ／ＦをＰＣＩバス等のホストの内部バスと、ＳＣＳＩバス等のホスト計算機とデ

ィスクアレイ装置をケーブルで接続する外部バスの両者を選択的に使用することができる。

【０２４０】また、ディスクアレイ装置２のバックボード２９０とディスクアレイコントローラ３を接続するコネクタを設けることで、両者を接続した際には自動的にホストＩ／ＦとしてＳＣＳＩを選択するようにできる。

【０２４１】本実施形態によれば、ディスクアレイコントローラに低価格なＭＰＵを使用したディスクアレイ制御システムにおいて、ＭＰＵの能力で性能が制限されることなく高性能なディスクアレイ制御が実現できる。

【０２４２】また、内蔵型ハイブリットアレイ構成のディスクアレイ制御システムと、外付け型ハイブリットアレイ構成のディスクアレイ制御システムとを同一のディスクアレイコントローラを用いて選択的に構成することができる。

【０２４３】次に、図２３～図２４を用いて、本発明の第３の実施形態によるディスクアレイ制御システムについて説明する。上述した第２の実施形態においては、ハイブリットアレイ構成のディスクアレイ制御システムについて説明したが、本実施形態においては、ディスクアレイコントローラ３を、「ハードアレイ」としても、「ハイブリッドアレイ」としても動作可能とし、この両者を選択的に使用できるよう構成したものである。

【０２４４】ここで、「ハードアレイ」とは、ディスクアレイコントローラ３においてすべてのディスクアレイ制御を行うもので、本明細書においては特にディスクアレイへの論理アドレスを各ディスクのディスクアドレスへアドレス変換する手段が、ディスクアレイコントローラ３上にあるディスクアレイの制御方法と定義する。

【０２４５】また、「ハイブリッドアレイ」とは、ディスクアレイコントローラ３とホスト計算機１の両方で分担してディスクアレイ制御を行うもので、本明細書においては特に上記アドレス変換する手段がホスト計算機１上にあり、さらに、ディスクアレイ制御のアドレス変換以外のある部分がディスクアレイコントローラ３上にある、ディスクアレイの制御方法と定義する。

【０２４６】すなわち、「ハイブリッドアレイ」は、上述した第１，２の実施形態のように、上位ディスクアレイ制御手段（上位ディスクアレイ制御ドライバ）がホスト計算機１上にあるもので、「ハードアレイ」は、上位ディスクアレイ制御部がディスクアレイコントローラ３上にあるものである。

【０２４７】最初に、図２３を用いて、本発明の第３の実施形態によるディスクアレイ制御システムのソフトウエアの全体構成について説明する。図２３は、本発明の第３の実施形態によるディスクアレイ制御システムのソフトウエアの全体構成のブロック図である。なお、図２と同一符号は、同一部分を示しており、以下の説明においては、図２との相違点を中心に説明する。

【０２４８】図２３において、ホスト計算機１は、図２に示すプログラムに加えて、さらに、ＣＰＵ１０で動作するディスクドライバ１０６を備えている。また、ディスクアレイコントローラ３のＭＰＵ３０は、図２に示すプログラムに加えて、さらに、第２上位ディスクアレイ制御部３０４と、ディスクアレイモード制御部３０６を備えている。

【０２４９】本実施形態におけるハード構成としては、図１若しくは図２０に示したハード構成とすることができる。本実施形態においては、図２０に示したハード構成に加えて、さらに、ディスクアレイコントローラは、ディスクアレイモード選択手段４４を備え、ディスクアレイ装置は、ハードアレイ選択手段２９を備えている。ハードアレイ選択手段２９は、図２０のＳＣＳＩ選択手段２５と全く同一に構成することができる。

【０２５０】ディスクアレイモード選択手段４４とハードアレイ選択手段２９の接続は、図２１に示したように、バックボードＩ／Ｆコネクタ３５３，２５１により行うことができる。

【０２５１】ディスクドライバ１０６は、ホスト計算機１に単体のディスクを接続するときに、ファイルシステムからのディスクアクセス要求をディスクへのコマンドに変換するものである。一般に、ＯＳ１０１は、標準でディスクドライバ１０６を備えている。

【０２５２】第２上位ディスクアレイ制御部３０４は、ファイルシステム１０２の代わりにＰＣＩ　Ｉ／Ｆ制御部３０１やＳＣＳＩ制御部３０２からディスクアレイへの論理コマンドを受信するものであり、この機能においては、上位ディスクアレイ制御ドライバ１０３の機能と同一である。

【０２５３】ディスクアレイモード選択手段４４の構成は、図３もしくは図２２に示したホストＩ／Ｆモード選択手段と全く同様に構成できる。本実施形態においては、図２２に示したホストＩ／Ｆモード選択手段と同様に構成したものとする。

【０２５４】また、ディスクアレイモード選択手段４４は、ディスクアレイモードとして、「ハイブリッドアレイモード」と「ハードアレイモード」のいずれを選択するかのＤｉｓｋＡｒｒａｙ＿Ｍｏｄｅ信号をＭＰＵ３０に送出する。

【０２５５】ここで、図２４を用いて、ディスクアレイモード選択手段４４が送出するＤｉｓｋＡｒｒａｙ＿Ｍｏｄｅ信号について説明する。図２４は、本発明の第３の実施形態によるディスクアレイ制御システムにおけるディスクアレイコントローラのディスクアレイモード選択手段が送出するＤｉｓｋＡｒｒａｙ＿Ｍｏｄｅ信号の論理図である。

【０２５６】図２４に示すように、ＤｉｓｋＡｒｒａｙ＿Ｍｏｄｅ信号が”１”の時には、ディスクアレイモードとして、「ハイブリッドアレイモード」を選択し、ＤｉｓｋＡｒｒａｙ＿Ｍｏｄｅ信号が”０”の時には、デ

ィスクアレイモードとして、「ハードアレイモード」のいずれを選択するようにしている。なお、ＤｉｓｋＡｒｒａｙ＿Ｍｏｄｅ信号［１：０］は、本発明の第５の実施形態に関するものであり、この点については、後述する。

【０２５７】ディスクアレイモード制御部３０６は、ディスクアレイモード選択手段４４が送出するＤｉｓｋＡｒｒａｙ＿Ｍｏｄｅ信号に従い、「ハードアレイモード」、「ハイブリッドアレイモード」のどちらでディスクアレイコントローラ３を動作させるかを決定する。

【０２５８】「ハイブリッドアレイモード」の時には、ディスクアレイモード制御部３０６は、第２上位ディスクアレイ制御部３０４をディスエーブルにする。また、ディスクアレイモード制御部３０６は、下位ディスクアレイ制御部３０３がコントローラコマンドをＰＣＩもしくはＳＣＳＩのホストＩ／Ｆ制御部（ＰＣＩＩ／Ｆ制御部３０１，ＳＣＳＩ制御部３０２）から受信するように制御する。

【０２５９】「ハードアレイモード」の時には、ディスクアレイモード制御部３０６は、第２上位ディスクアレイ制御部３０４をイネーブルにする。また、ディスクアレイモード制御部３０６は、下位ディスクアレイ制御部３０３がコントローラコマンドを第２上位ディスクアレイ制御部３０４から受信するように制御する。

【０２６０】ホストＩ／Ｆ３２，３３は、「ハイブリッドモード」，「ハードアレイモード」のいずれが選択されていても、ＰＣＩ　Ｉ／Ｆ３２とＳＣＳＩ３３のどちらでも使用できる。即ち、ＰＣＩ　Ｉ／Ｆ３２を使用することにより、図１に示したような「内蔵型」の構成とすることができ、また、ＳＣＳＩ３３を使用することにより、図２０に示したような「内蔵型」の構成とすることができる。

【０２６１】このようにホストＩ／Ｆモードとディスクアレイモードは独立なので、上記のようにディスクアレイモード選択手段４４を構成した場合、バックボードＩ／Ｆコネクタ３５３，２５１は、それぞれに専用な信号線を有する必要がある。

【０２６２】また、図３に示したホストＩ／Ｆモード選択手段３７のように、スイッチを用いてディスクアレイモード選択手段４４を構成した場合、スイッチはそれぞれ独立に備える必要がある。

【０２６３】本実施形態においては、図２０に示したように、ホストＩ／ＦとしてＳＣＳＩ３３が選択されていて、ディスクアレイコントローラ３がディスクアレイコントローラ３のバックボード２９０に接続しているものとする。ディスクアレイ装置２は、ハードアレイ選択手段２９を備えるので、図２２に示したＳＣＳＩ選択手段によるホストＩ／Ｆの選択と同様に、ハードアレイ選択手段２９により「ハードアレイモード」が選択される。もちろん、ハードアレイ選択手段を切り換え可能に構成し、「ハードアレイモード」と「ハイブリッドアレイモード」を切り換えることもできる。

【０２６４】上述したように、「ハードアレイ」と「ハイブリッドアレイ」の何れのディスクアレイでも、同一ディスクアレイコントローラを用い、かつ、ホストＩ／ＦもＰＣＩやＳＣＳＩによらず構築することができ、高性能が必要な場合には高速なホスト計算機のＣＰＵで上位ディスクアレイ制御を行える「ハイブリッドアレイモード」を選択でき、また、ホスト計算機やＯＳに非依存なディスクアレイを構築したい際や、ＣＰＵの負担を下げたい（ＣＰＵ負荷率を低減したい）際には、標準のディスクドライバを使用できる「ハードアレイ」を選択することができるものである。このように、構築自由度の高いディスクアレイを構成することができるものである。

【０２６５】以上説明したように、本実施形態によれば、ホスト計算機のホストバスやＯＳや、動作させるアプリケーションの要求性能に応じて、「ハードアレイ」方式のディスクアレイと、「ハイブリッドアレイ」方式のディスクアレイを、選択的に切り換えることができるディスクアレイコントローラを実現できるものである。

【０２６６】次に、図２５～図２６を用いて、本発明の第４の実施形態によるディスクアレイ制御システムについて説明する。従来のホスト計算機の内部バスに直結するタイプのディスクアレイ装置は、複数のホスト計算機により共用できないものであったが、本実施形態においては、これを可能とするものである。

【０２６７】最初に、図２５を用いて、本発明の第４の実施形態によるマルチアレイ式のハイブリッドアレイ構成のディスクアレイ制御システムの全体構成について説明する。図２５は、本発明の第４の実施形態によるディスクアレイ制御システムのハードウエアの全体構成のブロック図である。なお、図１及び図２０と同一符号は、同一部分を示しており、以下の説明においては、図１との相違点を中心に説明する。

【０２６８】図２５に示すように、本実施形態においては、２台のホスト計算機１ａ、１ｂを備えている。ホスト計算機１ａは、図１に示したホスト計算機１と同一である。また、ホスト計算機１ｂは、図２０に示したホスト計算機１と同一である。ホスト計算機１ａは、ディスクアレイコントローラ３をＰＣＩ　Ｉ／Ｆ３２で接続しており、ディスクアレイ装置２を内蔵している内蔵型の構成となっている。また、ホスト計算機１ｂは、ディスクアレイコントローラ３とＳＣＳＩ３３で接続している外付け型の構成となっている。

【０２６９】ディスクアレイコントローラ３のホストＩ／Ｆ選択手段３７は、ＩＦ＿Ｍｏｄｅ［１：０］信号を送出する。ここで、図２６を用いて、ホストＩ／Ｆ選択手段３７が送出するＩＦ＿Ｍｏｄｅ［１：０］信号について説明する。図２６は、本発明の第４の実施形態によ

るディスクアレイ制御システムにおけるディスクアレイコントローラのホストＩ／Ｆ選択手段が送出するＩＦ＿Ｍｏｄｅ［１：０］信号の論理図である。

【０２７０】ホストＩ／Ｆ選択手段３７は、図２６に示すようにＩＦ＿Ｍｏｄｅ［１：０］信号をＩＦ＿Ｍｏｄｅ０とＩＦ＿Ｍｏｄｅ１の２本出力する。そして、ＩＦ＿Ｍｏｄｅ［１：０］信号＝（０，１）の状態では、ホストＩ／Ｆとして、ＰＣＩＩ／Ｆ３２が選択され、ＩＦ＿Ｍｏｄｅ［１：０］信号＝（１，０）の状態では、ホストＩ／Ｆとして、ＳＣＳＩ３３が選択され、ＩＦ＿Ｍｏｄｅ［１：０］信号＝（１，１）の状態では、ホストＩ／Ｆとして、ＰＣＩＩ／Ｆ３２とＳＣＳＩ３３との両方が選択されるクラスタモードとなるものである。即ち、ディスクアレイコントローラ３が備えている２つのホストＩ／Ｆの何れか一方，若しくは両方を使用可能にする。何れか一方を使用する際には、信号線が２本あることをのぞき、図１若しくは図２０に示した実施形態と同様である。

【０２７１】また、両方使用する際には、ホストＩ／Ｆモード制御部３０５は、ＩＦ＿Ｍｏｄｅ［１：０］＝（１、１）の状態で信号を受信すると、初期化部３０３５に通知し、初期化部３０３５はいずれのホストＩ／Ｆからのコントローラコマンドも処理できるようにＰＣＩ　Ｉ／Ｆ制御部３０１，およびＳＣＳＩ制御部３０２を設定する。

【０２７２】また、ディスクアレイモード選択手段４４により決定されたディスクアレイモードで動作しながら、両方のホストＩ／Ｆで受信したコントローラコマンドもしくは論理コマンドを処理するように、下位ディスクアレイ制御部３０３と第２上位ディスクアレイ制御部３０４を設定する。

【０２７３】この際、ディスクアレイモードに応じ、ホスト計算機は上位ディスクアレイ制御ドライバもしくはディスクドライバのいずれか一方を備える必要がある。

【０２７４】また、ディスクアレイモード選択手段４４の出力信号ＤｉｓｋＡｒｒａｙ＿Ｍｏｄｅ信号も２本に拡張することで、各ホストＩ／Ｆ毎にディスクアレイモードを設定することも可能である。すなわち、例えば、ホスト計算機１ａは、「ハイブリッドアレイモード」で動作し、ホスト計算機１ｂは、「ハードアレイモード」で動作することが可能である。

【０２７５】この際、ＰＣＩ　Ｉ／Ｆ制御部３０１は、直接下位ディスクアレイ制御部３０３に受信したコントローラコマンドを引き渡し、また、ＳＣＳＩ制御部３０２は、第２上位ディスクアレイ制御部３０４に受信した論理コマンドを引き渡し、第２上位ディスクアレイ制御部３０４が、下位ディスクアレイ制御部３０３にコントローラコマンドを引き渡す。

【０２７６】以上のように、一台のディスクアレイ装置を、１台のホスト計算機１ａに内蔵した状態でもう１台のホスト計算機１ｂを同一ディスクアレイ装置に接続することができるので、ディスクアレイを２台のホスト計算機で共用できるようになる。

【０２７７】また、このように１台のディスクアレイ装置を２台のホスト計算機で共用できるので、２台のホスト計算機１ａ，１ｂの何れか一方を通常は動作していないスタンバイ機として運用し、普段動作する現用機が故障等によりダウンしたさいにスタンバイ機を動作させる、スタンバイ構成や、または両方のホスト計算機を常に動作させるクラスタ構成に対応するディスクアレイを構成できるものである。

【０２７８】また、上述した例では、ホスト計算機が２台の場合で説明したが、ＳＣＳＩにさらに多くのホスト計算機を接続することで、ｎ台のホスト計算機で唯一のディスクアレイを共用することができる。

【０２７９】以上のように、本実施形態によれば、「内蔵型ディスクアレイ」を実現した時にも、クラスタ構成や、スタンバイ構成等の複数台のホスト計算機で１台のディスクアレイを共用できるディスクアレイを実現できるようになる。

【０２８０】次に、図２４及び図２５を用いて、本発明の第５の実施形態によるディスクアレイ制御システムについて説明する。図２５に示したディスクアレイコントローラ３を内蔵したホスト計算機１ａの全体を一つのディスクアレイ装置とみなした例について説明する。この場合、ホスト計算機１ａは、ホスト計算機としての機能は有しておらず、ユーザのアプリケーションプログラムや，ネットワークや，グラフィック処理等は行わないものである。

【０２８１】ここで、図２５に示したディスクアレイコントローラ３のディスクアレイモード選択手段４４を拡張し、図２４に示すように、ＤｉｓｋＡｒｒａｙ＿ＭｏｄｅをＤｉｓｋＡｒｒａｙ＿Ｍｏｄｅ［１：０］に２ビットに拡張することにより、「ハイブリッドアレイモード」及び「ハードアレイモード」に加えて、「超ディスクアレイモード」を設ける。この場合、ＳＣＳＩ３３のみがホストＩ／Ｆとなる。

【０２８２】ディスクアレイモード選択手段４４が送出するＤｉｓｋＡｒｒａｙ＿Ｍｏｄｅ［１：０］＝（１，１）の状態では、ディスクアレイモードとして、「超ディスクアレイモード」に初期設定する。

【０２８３】図２５に示したホスト計算機１ｂのＣＰＵ１０は、ディスクアレイコントローラ３に論理コマンドを発行する。図２３において説明したように、ＳＣＳＩ制御部３０２は、論理コマンドを受信し、メモリ４２に格納する。ＭＰＵ３０のＰＣＩ　Ｉ／Ｆ制御部３０１は、ＰＣＩ　Ｉ／Ｆコントローラ３２１のＤＭＡ制御部を起動し、ホスト計算機１ａの主記憶メモリ１１にＤＭＡ転送する。

【０２８４】転送が終了すると、ＰＣＩ　Ｉ／Ｆ制御部

はＤｏｏｒＢｅｌｌレジスタにステータスを設定し、ＣＰＵ１０に割り込みを発行する。ＣＰＵ１０のＰＣＩ Ｉ／Ｆドライバは割り込みを受信すると、ＤｏｏｒＢｅｌｌレジスタのステータスをリードし、論理コマンドが転送されたことを認識する。上位ディスクアレイ制御ドライバ１０３は、論理コマンドを取り出し、以下、図１に示した実施形態と同様に動作する。

【０２８５】論理コマンドの実行が終了すると、上位ディスクアレイ制御ドライバ１０３はＰＣＩ Ｉ／Ｆドライバ３２１を制御し、論理コマンド終了ステータスを、図１に示した実施形態のコマンドリンク発行パケットの発行同様、ディスクアレイコントローラ３のメモリ４２に転送する。ＭＰＵ３０のＳＣＳＩ制御部３０２は、図２３に示した実施形態と同様に、ホスト計算機１ｂに論理コマンドの終了を報告し、処理を終了する。

【０２８６】以上のように、ホスト計算機１ａのＣＰＵ１０は、上位ディスクアレイ制御ドライバを動作させることに専念できるので、さらに高性能化を図ることができる。また、より高度なディスクアレイ制御アルゴリズムを搭載し、処理をすることができるので一層の高性能化を図ることができる。

【０２８７】本実施形態によれば、ホスト計算機にディスクアレイを内蔵したときに、そのホスト計算機自体を一つのディスクアレイ装置として構成し、さらに高性能なディスクアレイを実現することができる。

【０２８８】

【発明の効果】本発明によれば、ディスクアレイコントローラに低価格なＭＰＵを使用したディスクアレイ制御システムにおいて、ＭＰＵの能力で性能が制限されることなく高性能なディスクアレイ制御が実現できる。

【０２８９】また、本発明によれば、ディスクアレイコントローラに低価格なＭＰＵを使用したディスクアレイ制御システムにおいて、ＭＰＵの能力で性能が制限されることなく高性能なディスクアレイ制御が実現できる。

【０２９０】さらに、本発明によれば、ホスト計算機のホストバスやＯＳや、動作させるアプリケーションの要求性能に応じて、「ハードアレイ」方式のディスクアレイと、「ハイブリッドアレイ」方式のディスクアレイを、選択的に切り換えることができるディスクアレイコントローラを実現できるものである。

【０２９１】また、さらに、本発明によれば、複数台のホスト計算機で１台の「内蔵型ディスクアレイ」を共用できるディスクアレイを実現でき、クラスタ構成やスタンバイ型構成を実現することができるものである。

【０２９２】さらに、本発明によれば、ホスト計算機にディスクアレイを内蔵したときに、そのホスト計算機自体を一つのディスクアレイ装置の一部として構成できるので、ホスト計算機の高性能なＣＰＵを利用した高性能なディスクアレイを実現できるものである。

【図面の簡単な説明】

【図１】本発明の一実施形態によるディスクアレイ制御システムのハードウエアの全体構成のブロック図である。

【図２】本発明の一実施形態によるディスクアレイ制御システムのソフトウエアの全体構成のブロック図である。

【図３】本発明の一実施形態によるディスクアレイ制御システムのホストＩ／Ｆモード選択手段の構成を示す回路図である。

【図４】本発明の一実施形態によるディスクアレイ制御システムのホストＩ／Ｆモード選択手段の論理図である。

【図５】本発明の一実施形態によるディスクアレイ制御システムにおけるディスクアレイコントローラ検出処理を説明するフローチャートである。

【図６】本発明の一実施形態によるディスクアレイ制御システムにおけるディスクアレイのリード処理の説明図である。

【図７】本発明の一実施形態によるディスクアレイ制御システムにおける上位ディスクアレイ制御ドライバのコマンドリンクの発行方法について説明するフローチャートである。

【図８】本発明の一実施形態によるディスクアレイ制御システムにおけるホスト計算機とディスクアレイコントローラの間のコマンドリンクの発行方法の説明図である。

【図９】本発明の一実施形態によるディスクアレイ制御システムにおけるＭＰＵのコマンドリンクの割り込み処理について説明するフローチャートである。

【図１０】本発明の一実施形態によるディスクアレイ制御システムにおけるＭＰＵのＰＣＩ Ｉ／Ｆ制御部の中のＤＭＡ制御部の処理について説明するフローチャートである。

【図１１】本発明の一実施形態によるディスクアレイ制御システムにおけるＭＰＵの下位ディスクアレイ制御部の処理について説明するフローチャートである。

【図１２】本発明の一実施形態によるディスクアレイ制御システムにおけるＭＰＵの下位ディスクアレイ制御部の処理について説明するフローチャートである。

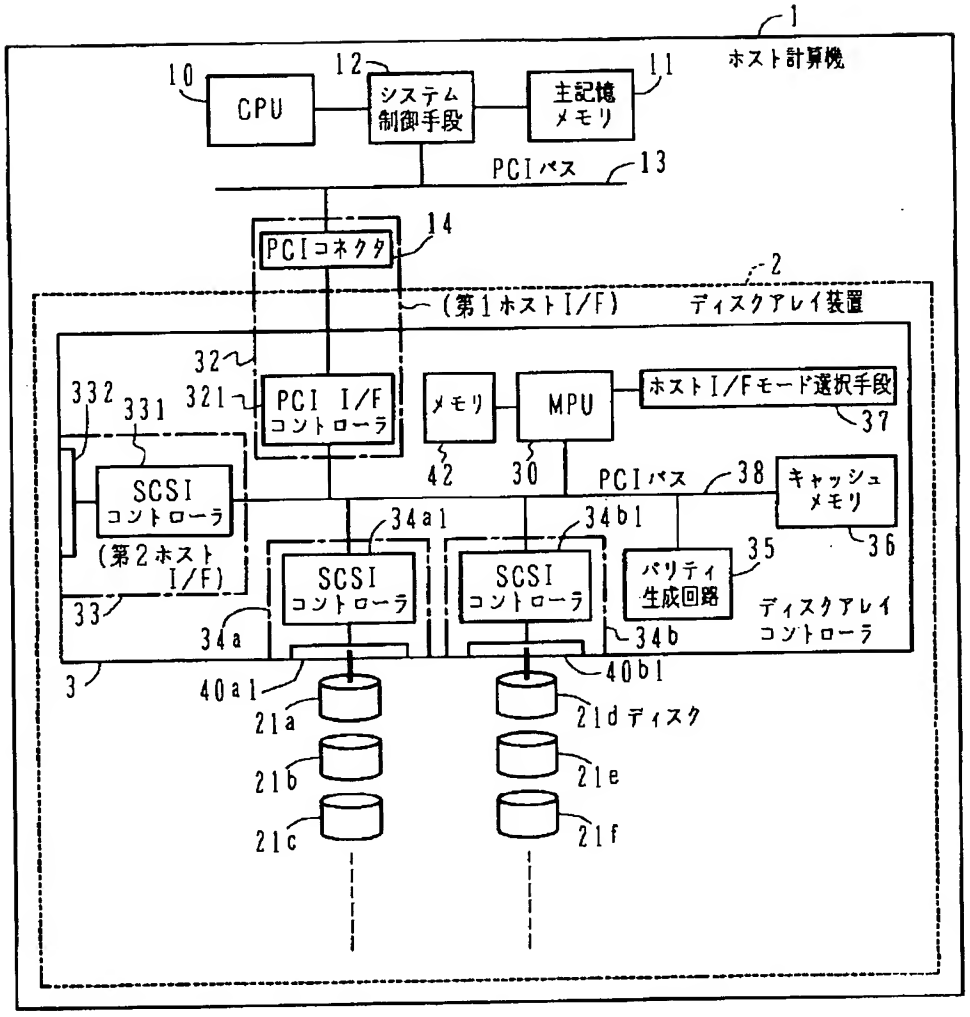
【図１３】本発明の一実施形態によるディスクアレイ制御システムにおけるＭＰＵの下位ディスクアレイ制御部のディスクコマンド制御部の処理について説明するフローチャートである。

【図１４】本発明の一実施形態によるディスクアレイ制御システムにおけるＭＰＵの下位ディスクアレイ制御部のディスクコマンド制御部の終了報告受信時の処理について説明するフローチャートである。

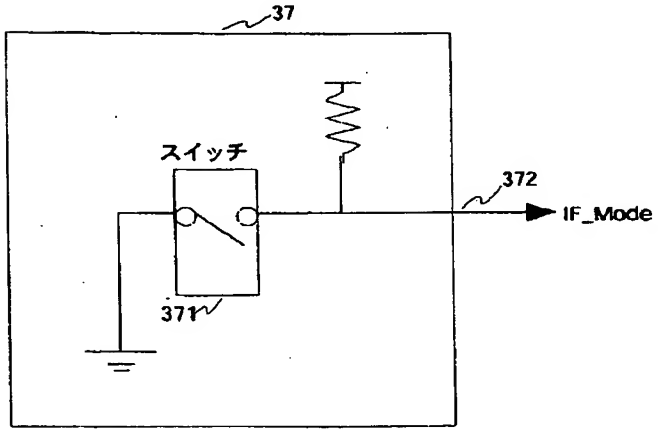
【図１５】本発明の一実施形態によるディスクアレイ制御システムにおけるホスト計算機のＰＣＩ Ｉ／Ｆドライバにおける終了報告処理について説明するフローチャ

ートである。
【図16】本発明の一実施形態によるディスクアレイ制御システムにおけるディスクアレイのライト処理の説明図である。
【図17】本発明の一実施形態によるディスクアレイ制御システムにおけるディスクアレイの第1のデステージ処理の説明図である。
【図18】本発明の一実施形態によるディスクアレイ制御システムにおけるディスクアレイの第2のデステージ処理の説明図である。
【図19】本発明の一実施形態によるディスクアレイ制御システムにおけるMPUの下位ディスクアレイ制御部のパリティ生成コマンド制御部の処理について説明するフローチャートである。
【図20】本発明の第2の実施形態によるディスクアレイ制御システムのハードウエアの全体構成のブロック図である。
【図21】本発明の第2の実施形態によるディスクアレイ制御システムのハードウエアのブロック斜視図である。
【図22】本発明の第2の実施形態によるディスクアレイ制御システムのホストI／Fモード選択手段の構成を示す回路図である。
【図23】本発明の第3の実施形態によるディスクアレイ制御システムのソフトウエアの全体構成のブロック図である。
【図24】本発明の第3の実施形態によるディスクアレイ制御システムにおけるディスクアレイコントローラのディスクアレイモード選択手段が送出するDiskArray_Mode信号の論理図である。
【図25】本発明の第4の実施形態によるディスクアレイ制御システムのハードウエアの全体構成のブロック図である。
【図26】本発明の第4の実施形態によるディスクアレイ制御システムにおけるディスクアレイコントローラのホストI／F選択手段が送出するIF_Mode［1：0］信号の論理図である。
【符号の説明】
1，1a，1b…ホスト計算機
10…CPU
101…OS
102…ファイルシステム
103…上位ディスクアレイ制御ドライバ
1031…論理コマンドアドレス変換部
1032…キャッシュ管理部
1034…コマンドI／F制御部
1035…コントローラ検出部
104…PCI I／Fドライバ
105…SCSIドライバ
106…ディスクドライバ
11…主記憶メモリ
12…システム制御手段
13…PCIバス
14…PCIコネクタ
15…SCSIコントローラ
16…SCSIコネクタ
17…SCSIケーブル
2…ディスクアレイ装置
21，21a，21b，21c，21d，21e，21f…ディスク
22…電源
23…クロック発生器
24…コネクタ
25…SCSI選択手段
26…SCSIコネクタ
28…筐体異常検出手段
29…ハードアレイ選択手段
3…ディスクアレイコントローラ
30…MPU
301…PCI I／F制御部
302…SCSI制御部
303…下位ディスクアレイ制御部
304…第2上位ディスクアレイ制御部
31…PCIカードエッジ
32…PCII／F
321…PCI I／Fコントローラ
3211…DMAC
33，34a，34b…SCSI
331…SCSIコントローラ
332…SCSIコネクタ
34a1，34b1…SCSIコントローラ
35…パリティ生成回路
36…キャッシュメモリ
37，37a…ホストI／Fモード選択手段
38…PCIバス
40a1，40b1…SCSIコネクタ
42…メモリ
43…筐体監視手段
44…ディスクアレイモード選択手段

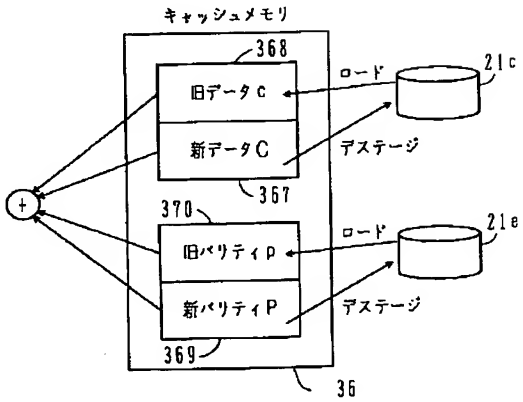
【図１】


1
ホスト計算機
10
CPU
12
システム制御手段
11
主記憶メモリ
PCIバス
13
14
PCIコネクタ
2
(第1ホストI/F)
ディスクアレイ装置
32
332
331
321
PCI I/F コントローラ
メモリ
MPU
ホストI/Fモード選択手段
37
SCSI コントローラ
(第2ホスト I/F)
42
30
PCIバス
38
キャッシュメモリ
34a1
34b1
SCSI コントローラ
SCSI コントローラ
パリティ生成回路
35
36
ディスクアレイ コントローラ
33
34a
34b
3
40a1
40b1
21a
21b
21c
21d ディスク
21e
21f

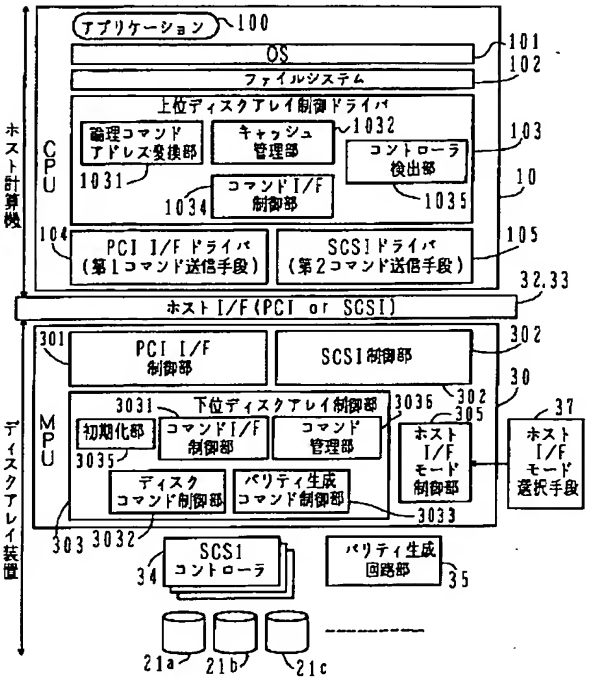
【図３】


37
スイッチ
372
IF_Mode
371

【図１７】


キャッシュメモリ
368
旧データC
新データC
ロード
デステージ
21c
367
370
旧パリティP
新パリティP
ロード
デステージ
21e
369
36

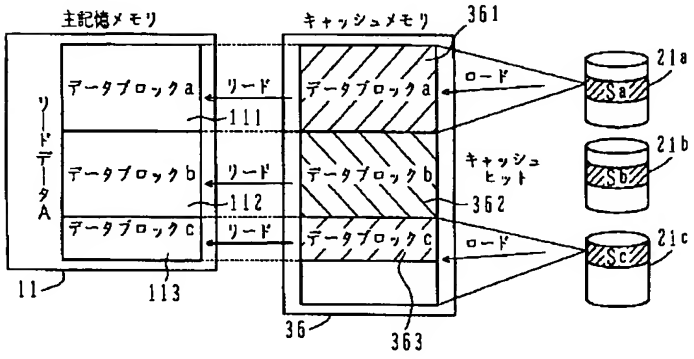
【図2】



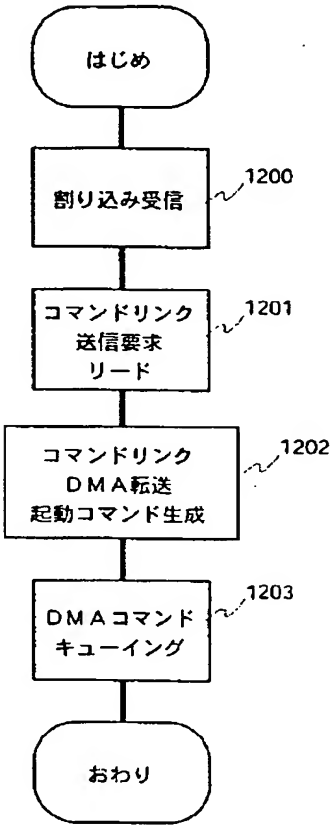
【図4】

IF_Mode	1	0
ホストI／F	PCI I／F（第一ホストI／F）	SCSI（第二ホストI／F）

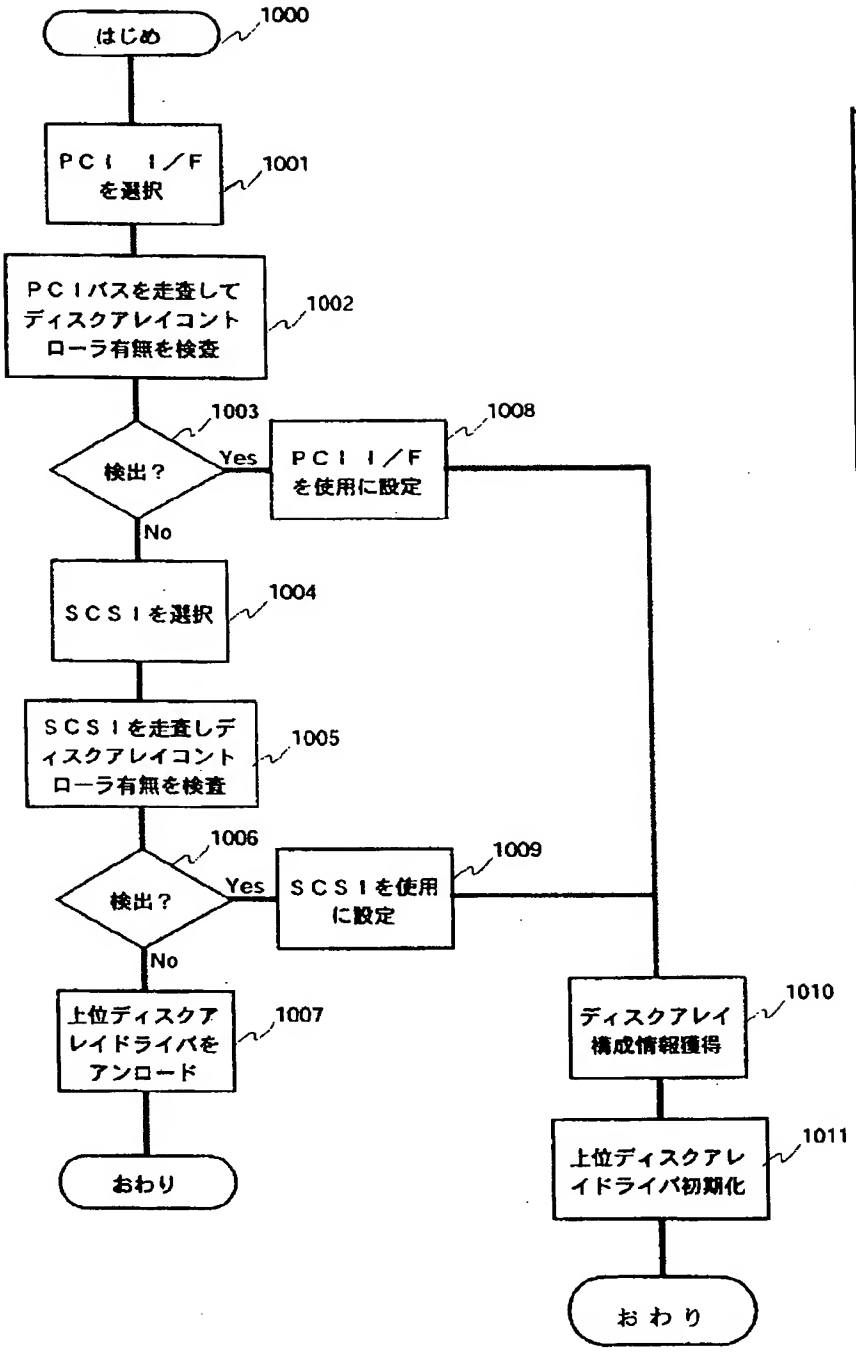
【図6】



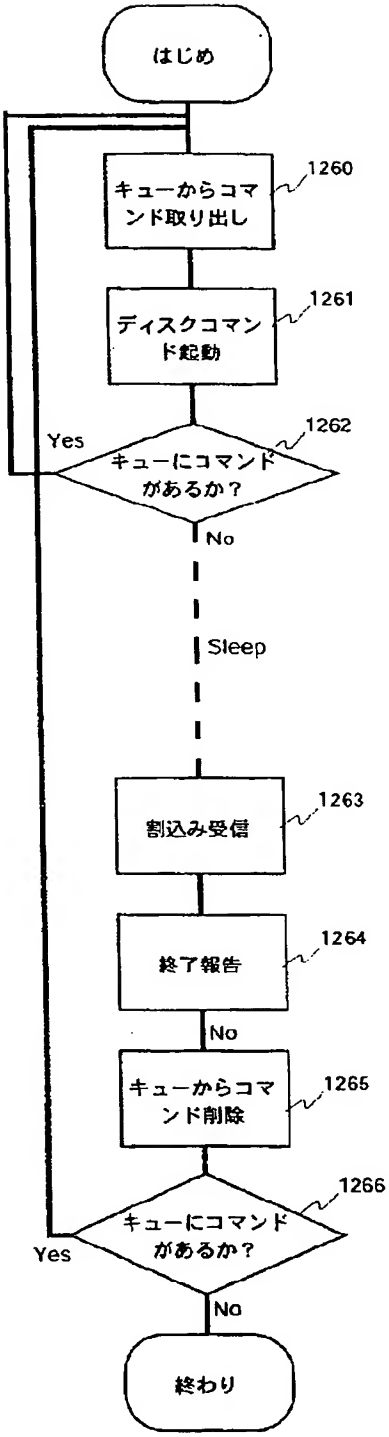
【図9】



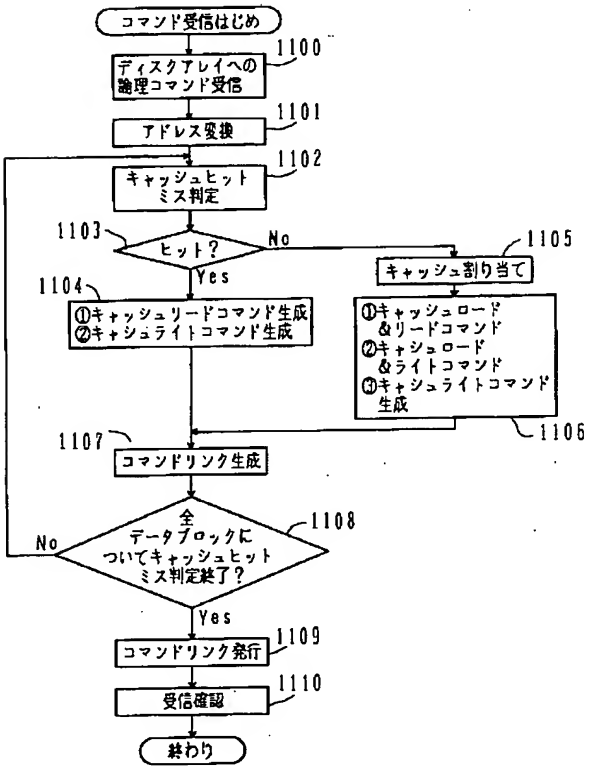
【図５】


はじめ
1000
PCI I/Fを選択
1001
PCIバスを走査してディスクアレイコントローラ有無を検査
1002
検出？
1003
Yes
PCI I/Fを使用に設定
1008
No
SCSIを選択
1004
SCSIを走査しディスクアレイコントローラ有無を検査
1005
検出？
1006
Yes
SCSIを使用に設定
1009
No
上位ディスクアレイドライバをアンロード
1007
おわり
ディスクアレイ構成情報獲得
1010
上位ディスクアレイドライバ初期化
1011
おわり

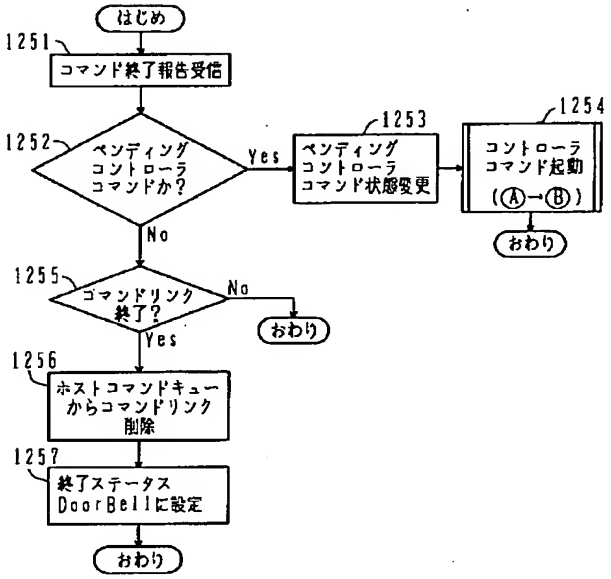
【図１３】


はじめ
キューからコマンド取り出し
1260
ディスクコマンド起動
1261
Yes
キューにコマンドがあるか？
1262
No
Sleep
割込み受信
1263
終了報告
1264
No
キューからコマンド削除
1265
キューにコマンドがあるか？
1266
Yes
No
終わり

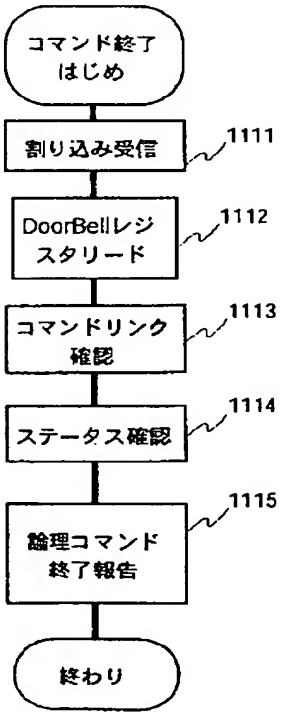
【図7】


コマンド受信はじめ
ディスクアレイへの論理コマンド受信
1100
アドレス変換
1101
キャッシュヒットミス判定
1102
ヒット？
1103
No
Yes
キャッシュ割り当て
1105
①キャッシュリードコマンド生成
②キャシュライトコマンド生成
1104
①キャッシュロード&リードコマンド
②キャシュロード&ライトコマンド
③キャシュライトコマンド生成
1106
コマンドリンク生成
1107
全データブロックについてキャッシュヒットミス判定終了？
1108
No
Yes
コマンドリンク発行
1109
受信確認
1110
終わり

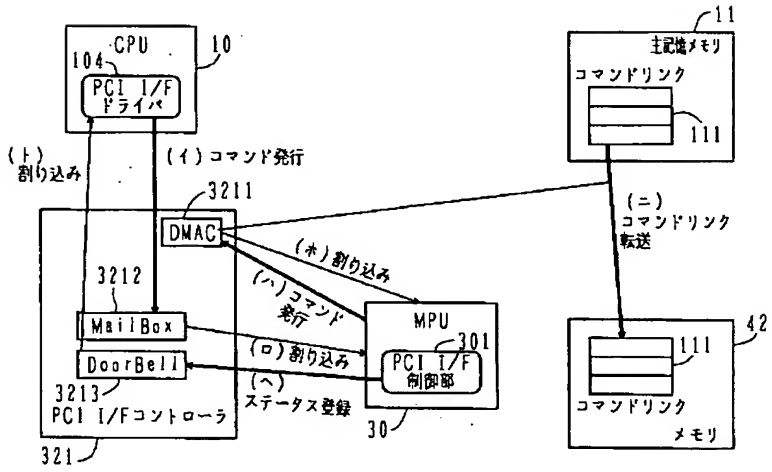
【図14】


はじめ
1251
コマンド終了報告受信
ペンディングコントローラコマンドか？
1252
Yes
No
ペンディングコントローラコマンド状態変更
1253
コントローラコマンド起動（Ⓐ→Ⓑ）
1254
おわり
コマンドリンク終了？
1255
No
Yes
おわり
ホストコマンドキューからコマンドリンク削除
1256
終了ステータスDoorBellに設定
1257
おわり

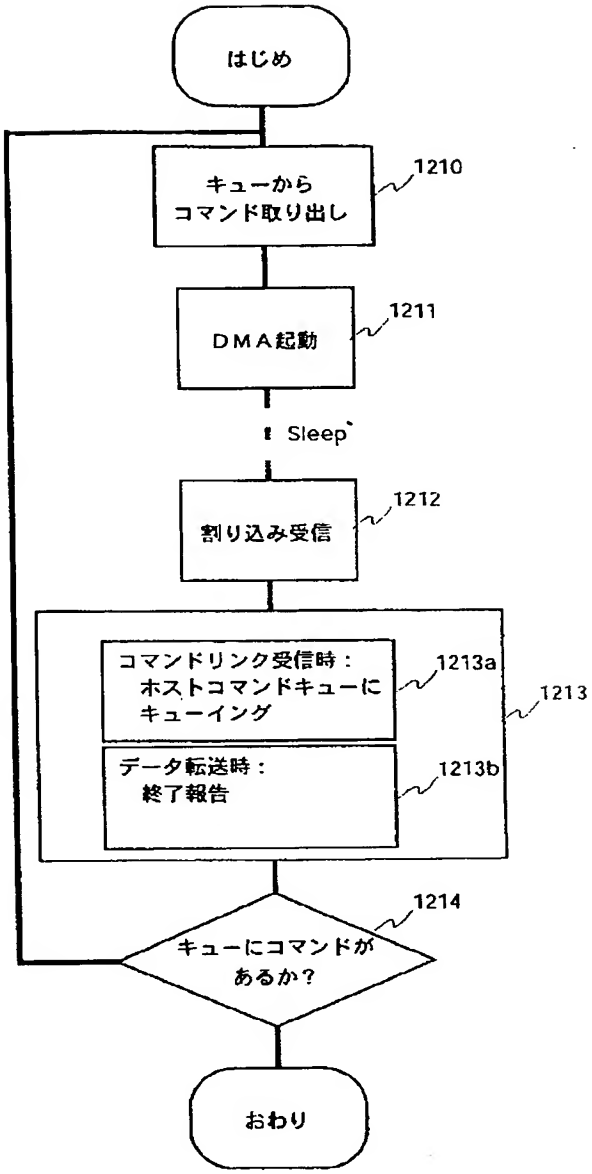
【図15】


コマンド終了はじめ
割り込み受信
1111
DoorBellレジスタリード
1112
コマンドリンク確認
1113
ステータス確認
1114
論理コマンド終了報告
1115
終わり

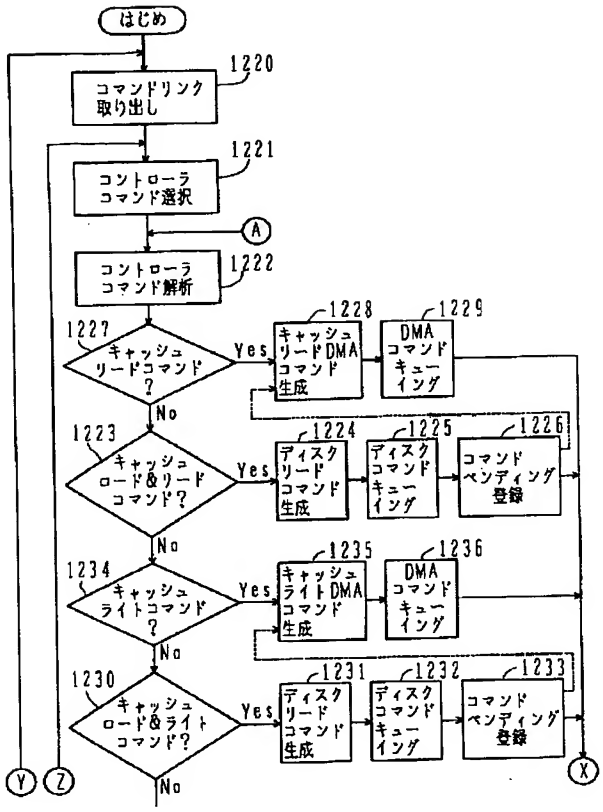
【図8】


CPU
10
104
PCI I/Fドライバ
(ト)割り込み
(イ)コマンド発行
3211
DMAC
(ホ)割り込み
(ハ)コマンド発行
3212
MailBox
DoorBell
(ロ)割り込み
(ヘ)ステータス登録
3213
PCI I/Fコントローラ
321
MPU
301
PCI I/F制御部
30
11
主記憶メモリ
コマンドリンク
111
(ニ)コマンドリンク転送
111
42
コマンドリンクメモリ

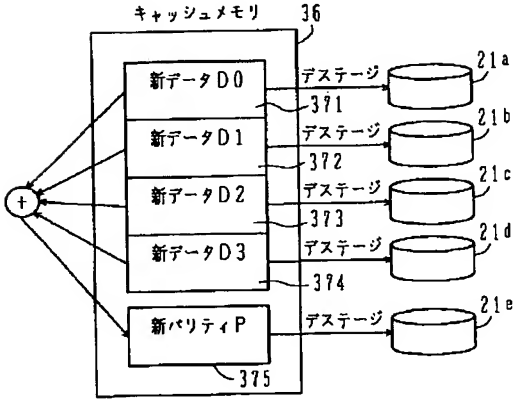
【図１０】



【図１１】



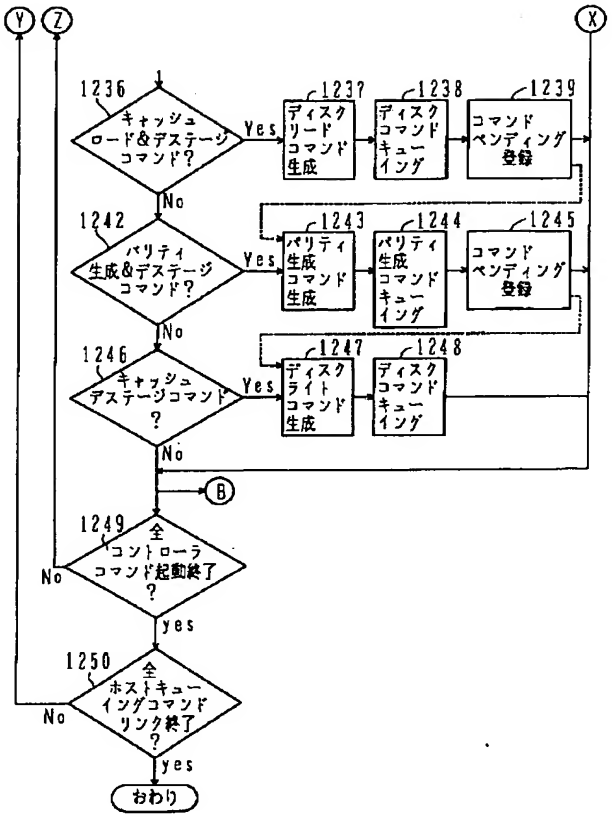
【図１８】



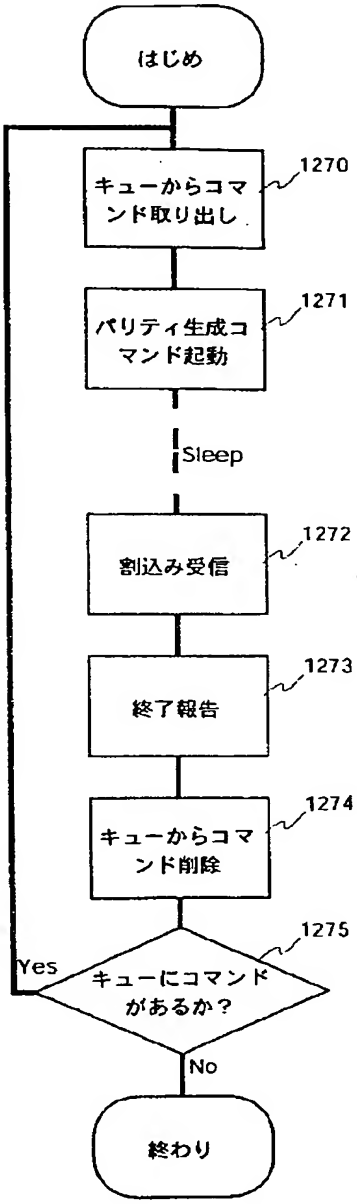
【図２４】

DiskArray_Mode	1	0	DiskArray_Mode[1:0] 1 0
ディスクアレイ モード	ハイブリッド アレイモード	ハードアレイ モード	無ディスクアレイ モード

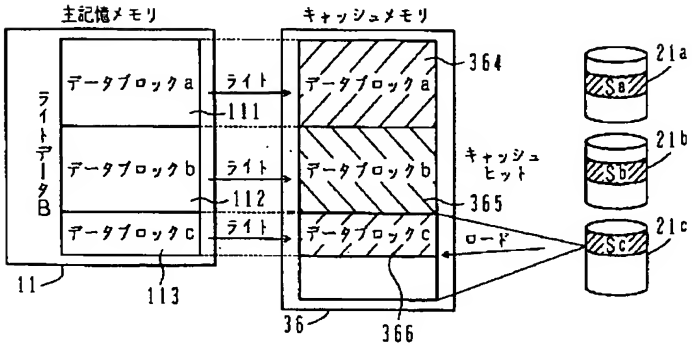
【図１２】


Y
Z
X
1236
キャッシュロード&デステージコマンド?
Yes
1237
ディスクリードコマンド生成
1238
ディスクコマンドキューイング
1239
コマンドペンディング登録
No
1242
パリティ生成&デステージコマンド?
Yes
1243
パリティ生成コマンド生成
1244
パリティ生成コマンドキューイング
1245
コマンドペンディング登録
No
1246
キャッシュデステージコマンド?
Yes
1247
ディスクライトコマンド生成
1248
ディスクコマンドキューイング
No
B
1249
全コントローラコマンド起動終了?
No
yes
1250
全ホストキューイングコマンドリンク終了?
No
yes
おわり

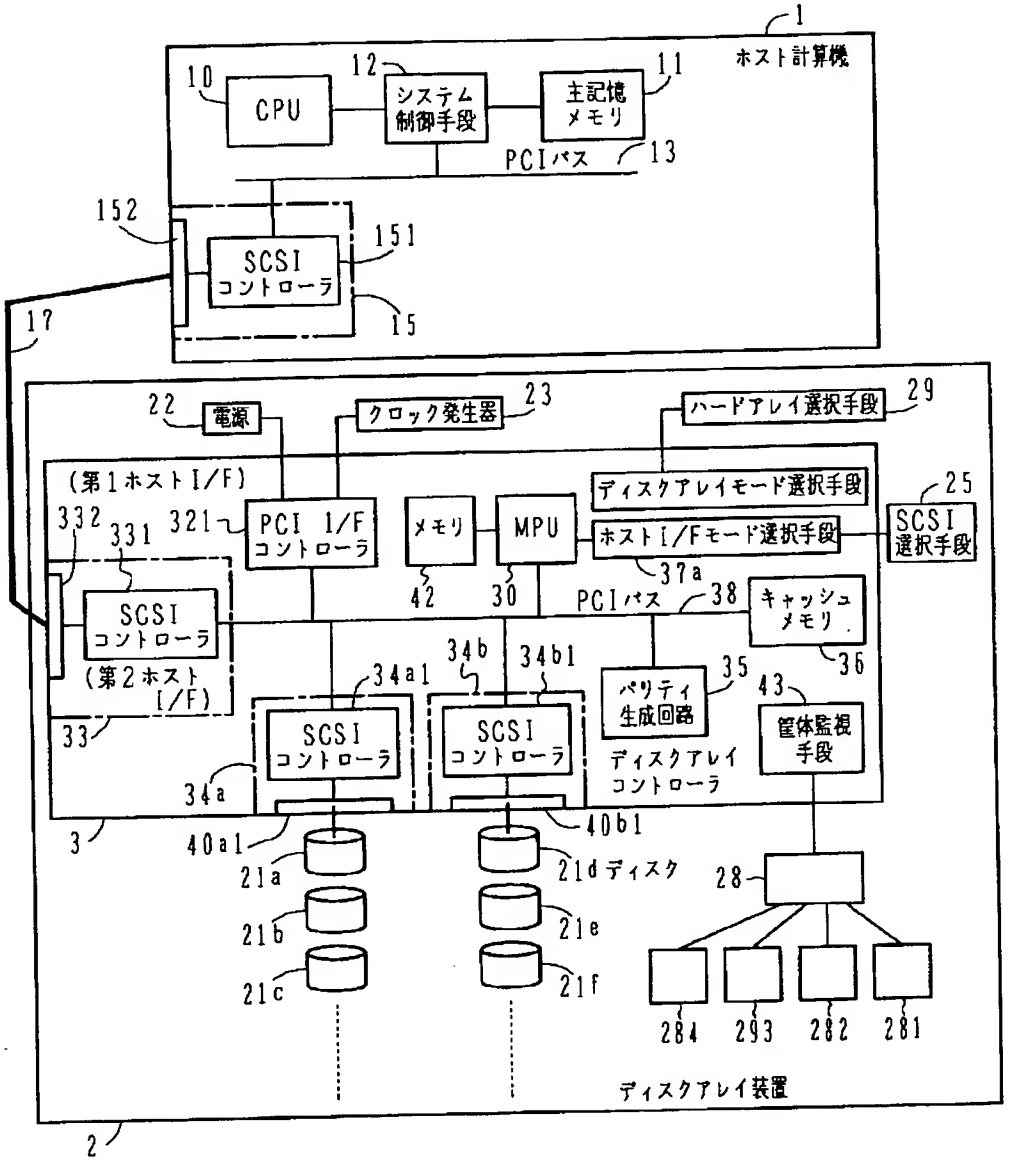
【図１９】


はじめ
1270
キューからコマンド取り出し
1271
パリティ生成コマンド起動
Sleep
1272
割込み受信
1273
終了報告
1274
キューからコマンド削除
1275
キューにコマンドがあるか?
Yes
No
終わり

【図１６】


主記憶メモリ
キャッシュメモリ
ライトデータB
データブロックa
データブロックb
データブロックc
ライト
111
112
113
11
364
365
366
36
キャッシュヒット
ロード
21a
21b
21c
Sa
Sb
Sc

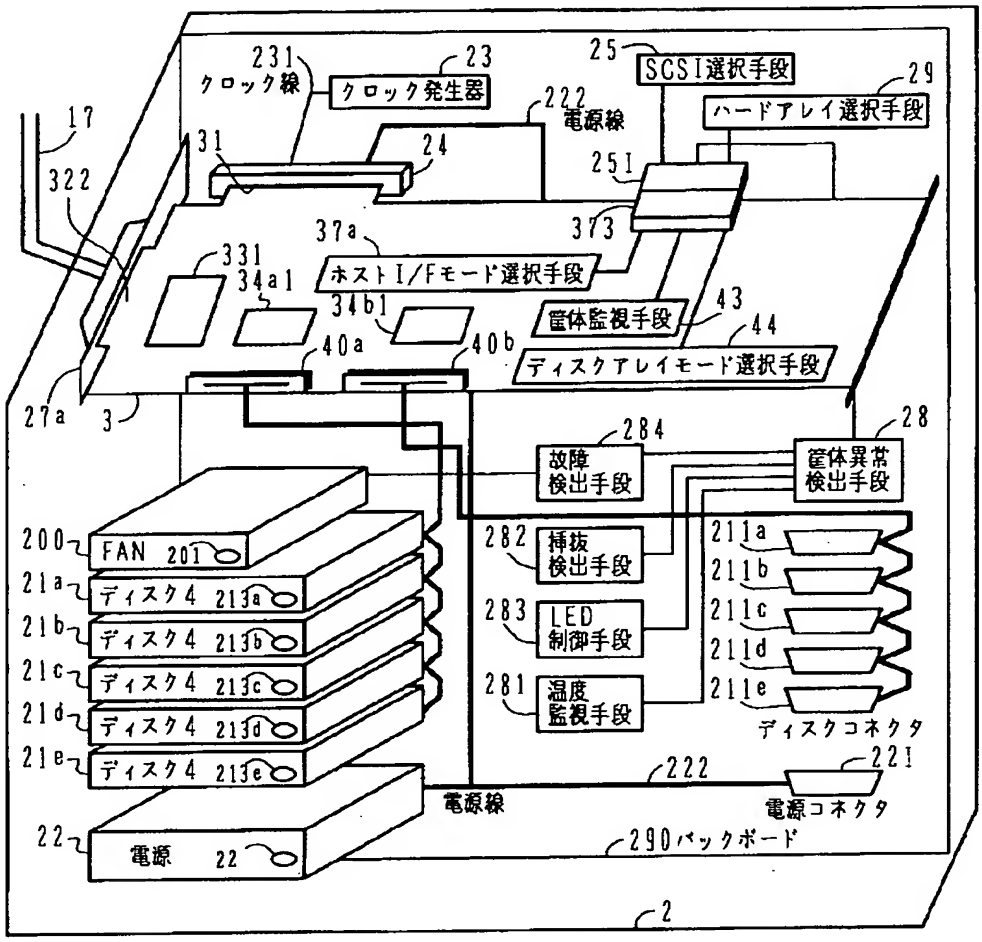
【図２０】



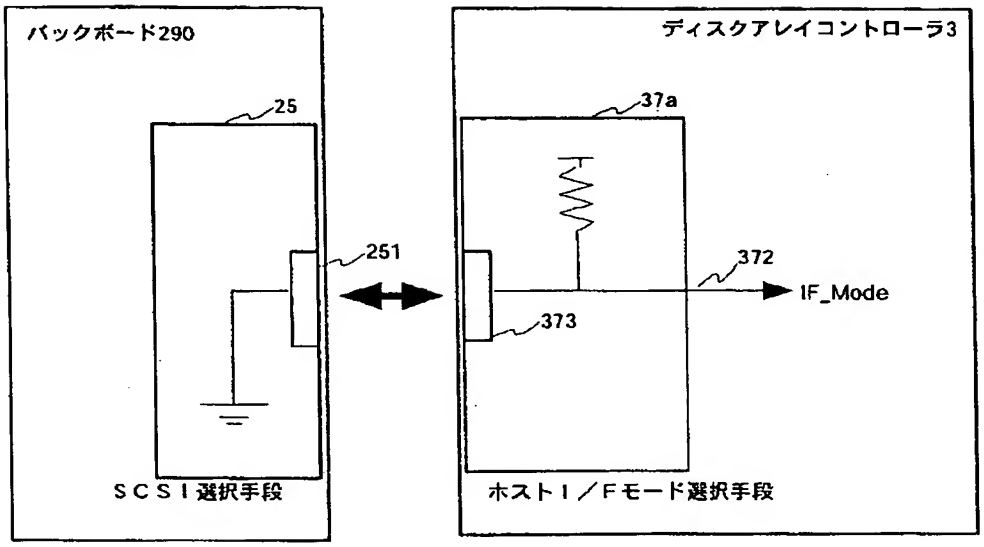
【図２６】

IF_Mode[1:0]	１１	１０	０１	００
ホストＩ／Ｆモード	両方使用 （クラスタモード）	ＳＣＳＩ使用	ＰＣＩ Ｉ／Ｆ使用	未使用

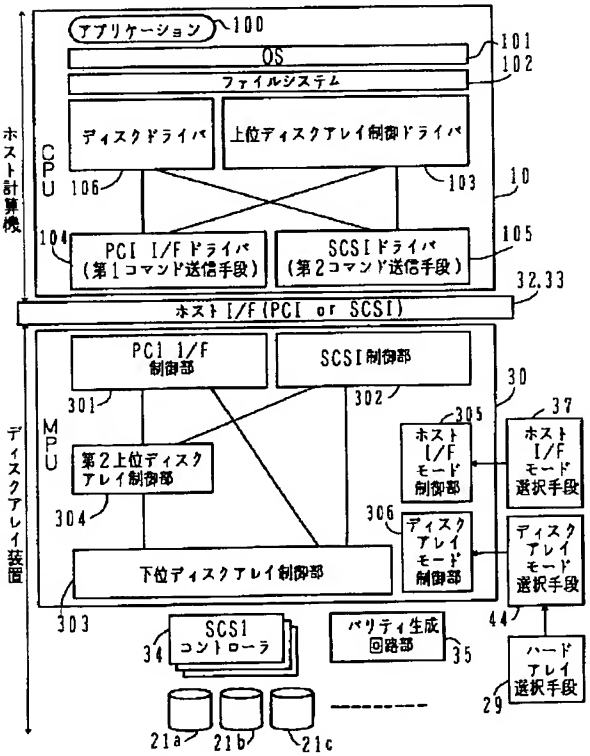
【図21】


231
クロック線
23
クロック発生器
25
SCSI選択手段
29
ハードアレイ選択手段
222
電源線
17
322
31
24
251
373
37a
ホストI/Fモード選択手段
331
34a1
34b1
筐体監視手段
43
44
ディスクアレイモード選択手段
40a
40b
27a
3
284
故障
検出手段
28
筐体異常
検出手段
200
FAN 201
21a
ディスク4 213a
21b
ディスク4 213b
21c
ディスク4 213c
21d
ディスク4 213d
21e
ディスク4 213e
282
挿抜
検出手段
283
LED
制御手段
281
温度
監視手段
211a
211b
211c
211d
211e
ディスクコネクタ
222
221
電源線
電源コネクタ
22
電源 22
290バックボード
2

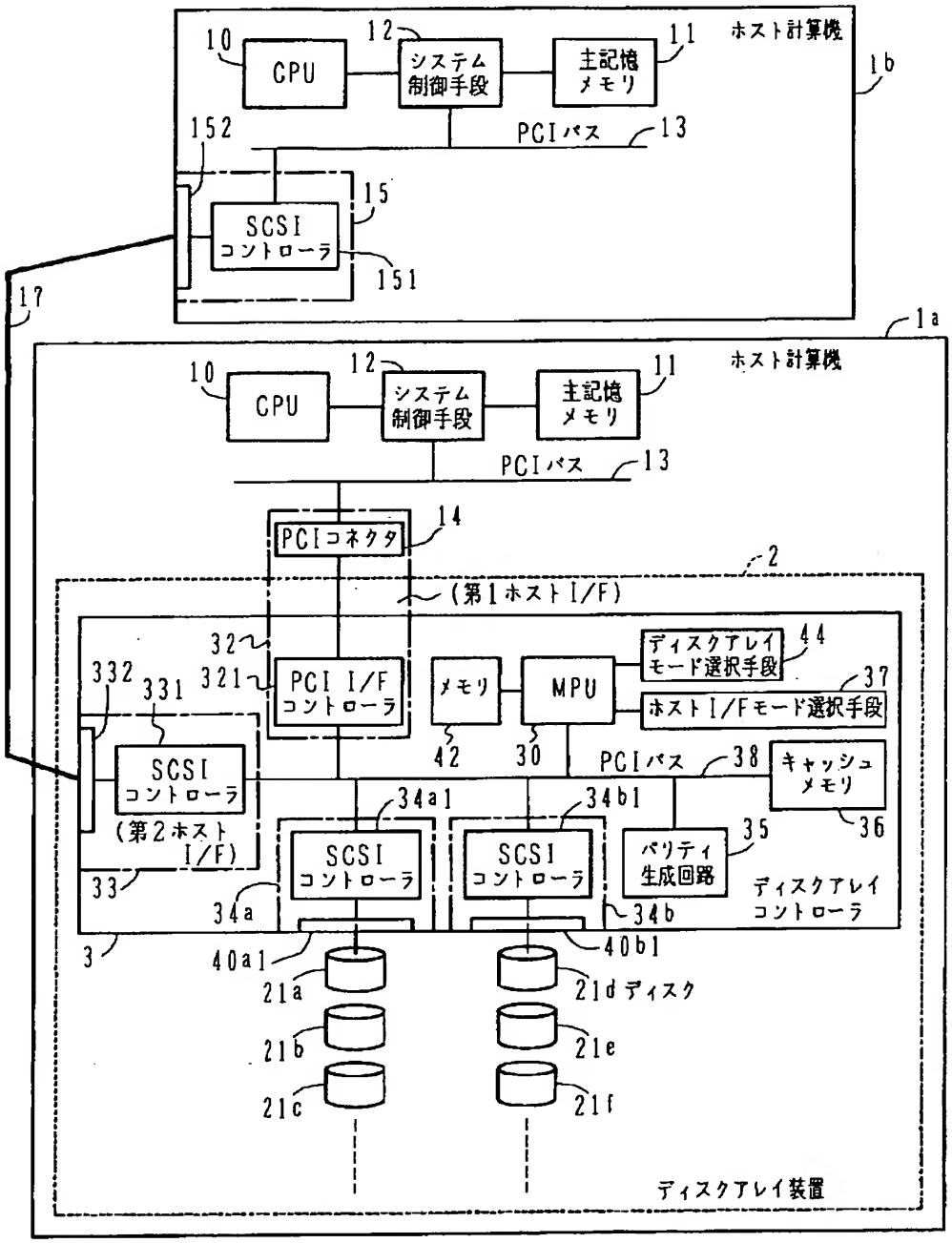
【図２２】


バックボード290
25
251
ＳＣＳＩ選択手段
ディスクアレイコントローラ3
37a
372
IF_Mode
373
ホストＩ／Ｆモード選択手段

【図２３】


ホスト計算機
アプリケーション
100
OS
101
ファイルシステム
102
CPU
ディスクドライバ
106
上位ディスクアレイ制御ドライバ
103
10
104
PCI I/F ドライバ
（第1コマンド送信手段）
SCSIドライバ
（第2コマンド送信手段）
105
32,33
ホストI/F(PCI or SCSI)
ディスクアレイ装置
PCI I/F
制御部
301
SCSI制御部
302
30
MPU
第2上位ディスク
アレイ制御部
304
下位ディスクアレイ制御部
303
305
ホスト
I/F
モード
制御部
306
ディスク
アレイ
モード
制御部
37
ホスト
I/F
モード
選択手段
ディスク
アレイ
モード
選択手段
44
ハード
アレイ
選択手段
29
SCSI
コントローラ
34
パリティ生成
回路部
35
21a
21b
21c

【図25】



フロントページの続き

(72)発明者 荒川 敬史
神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内

(72)発明者 八木沢 育哉
神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内

(72)発明者　高野　雅弘
神奈川県小田原市国府津2880番地　株式会
社日立製作所ストレージシステム事業部内